12-2021

Site-specific Effects of Lysine Acetylation on Aminoacyl-tRNA Synthetase

Hao Chen
University of Arkansas, Fayetteville

Site-specific Effects of Lysine Acetylation on Aminoacyl-tRNA Synthetase

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Cell and Molecular Biology

by

Hao Chen
Wenzhou Medical University
Bachelor of Science in Biotechnology, 2014
Wenzhou Medical University
Master of Science in Biology, 2017

December 2021
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

_____
Chenguang Fan, Ph.D.
Dissertation Director

_____
Roger Koeppe, Ph.D.
Committee Member

_____
Josh Sakon, Ph.D.
Committee Member

_____
Jeffrey Lewis, Ph.D.
Committee Member

Abstract

Aminoacyl-tRNA synthetases (AARSs) are an ancient and highly conserved family of enzymes which can catalyze a two-steps aminoacylation reaction to charge tRNAs with their cognate amino acids, thus playing crucial roles in ribosomal protein synthesis. Naturally, the accurate amino acids and tRNA recognition of these synthetases are essential to the fidelity of translation process. To assure the correct recognition, some of these synthetases have evolved with an editing function to help remove the mischarged tRNAs. In addition to these functions, AARSs are also involved in various biological processes ranging from transcription to translation. Currently, a series of proteomic studies have shown that AARSs are one of the enzymes with the most abundant acetylation. And a few studies have shown site-specific acetylation on AARSs can cause significant changes in enzyme activities. It is necessary to conduct more investigation on the relationship between acetylation and the function of AARSs.

However, due to the dynamic, reversible, transient, and uncontrollable pattern of lysine acetylation, it is usually challenging to learn its effects on proteins. To solve this problem, genetic code expansion strategy can be an ideal solution. This strategy offers a co-translational incorporation system (a pair of engineered orthogonal AARS/tRNA and an unassigned codon is required) which is able to introduce noncanonical amino acids including acetyllysine (AcK) into specific positions of object proteins.

Using an engineered pyrrolysyl-tRNA synthetase variant specific for AcK and its orthogonal $tRNA^{pyl}$, this study established an AcK incorporation system which could stably produce site-specific acetylated AARSs. Four AARSs in *Escherichia coli*, including class I

cysteinyl-tRNA synthetase (CysRS) as well as class II aspartyl-tRNA synthetase (AspRS), histidyl-tRNA synthetase (HisRS), and threonyl-tRNA synthetase (ThrRS), were chosen as objects in this study. The results indicate that the impacts of lysine acetylation could be different between two classes, and even within the class. In addition, this study also found that acetylation of ThrRS (at K169) could affect its editing function can cause mistranslation. To furtherly detect the possible mistranslation rate of ThrRS with site-specific acetylation in its editing domain, we also designed a GFP variant with only one Threonine (Thr) in position 203 which determines the fluorescence signal of GFP. Briefly, if the mistranslation occurs, then the Thr would be mistranslated to serine, and then the GFP probe would lose its fluorescence signal. This probe was then applied to indicate that K169 acetylation could affect the editing function of ThrRS and cause mistranslation *in vivo*.

Acknowledgments

First, I would like to express my great gratitude to my advisor, Dr. Chenguang Fan for offering the chance to join his lab and always supporting me no matter the good time or bad time. His extraordinary personality and rigorous scientific research attitude inspired me a lot. I learned how to design and conduct studies by his patient and systematic guidance, and I also get to access so many cutting-edge technologies in his lab. It is my honor to work with him and I am sincerely thankful for his mentoring during my doctoral study. I also want to thank my committee members, Dr. Roger Koeppe, Dr. Josh Sakon, and Dr. Jeffrey Lewis for their supports and help.

Secondly, I would like to thank Cell and Molecular Biology program and Chemistry and Biochemistry department at University of Arkansas for offering both wonderful study and research environment.

Thirdly, I am very grateful to my lab mates, especially Suman Venkat who has become one of my best friends. She helps me a lot not only in lab works but also in my personal life. I also want to thank the undergraduates, Denver Hudson, Tony Wang, and Carson Ercanbrack who helped me when we were doing experiments together.

Additionally, I also want to show my sincere love to the city, Fayetteville in Arkansas state which offers a good and peaceful environment.

Finally, I am especially and extremely thankful to my parents, Yuquan Chen and Zhumei Lin. Without their unconditional supports and love, I will not be able to achieve all my accomplishments today.

Table of contents

List of figures

List of Tables

List of published paper

Chapter II: Chen, H., et al., *Site-Specifically Studying Lysine Acetylation of Aminoacyl-tRNA Synthetases.* ACS chemical biology, 2019. **14**(2): p. 288-295.	Published
Chapter III: Chen, H., et al., *A Synthetic Reporter for Probing Mistranslation in Living Cells.* Frontiers in bioengineering and biotechnology, 2020. **8**: p. 623.	Published

CHAPTER I: Introduction

1.1 Aminoacyl-tRNA synthetases (AARSs)

Aminoacyl-tRNA synthetases (AARSs) are an ancient family of enzymes which essentially function in protein synthesis[1]. Canonically, AARSs are responsible for the correct translation of three-letter genetic code in mRNA, and they can catalyze a two-step aminoacylation reaction where the amino acid is activated by ATP hydrolysis to form an intermediate, aminoacyl-AMP, and then the aminoacyl group is transferred to either 3'-OH or 2'-OH of A in CCA acceptor sequence of cognate tRNA to form aminoacyl-tRNA which will be delivered to the ribosome and participate in protein synthesis (Figure 1.1)[1, 2].

Currently, there are a total 23 AARSs found in organisms, including 21 for 20 proteinogenic amino acids (two for lysine) plus pyrrolysyl-tRNA synthetase (PylRS) and phosphoseryl-tRNA synthetase (SepRS)[3, 4]. These enzymes can be divided into two distinct classes mainly featured by their own class-specific catalytic domains. In class I AARSs (methionyl-tRNA synthetase (MetRS), valyl-tRNA synthetase (ValRS), leucyl-tRNA synthetase (LeuRS), isoleucyl-tRNA synthetase (IleRS), cysteinyl-tRNA synthetase (CysRS), arginyl-tRNA synthetase (ArgRS), glutamyl-tRNA synthetase (GluRS), glutaminyl-tRNA synthetase (GlnRS), lysyl-tRNA synthetase-I (LysRS-I), tyrosyl-tRNA synthetase (TyrRS) and tryptophanyl-tRNA synthetase (TrpRS)), the class-conserved motifs HIGH and KMSKS play important role in the interaction with substrate ATP. However, in class II AARS (seryl-tRNA synthetase (SerRS), threonyl-tRNA synthetase (ThrRS), alanyl-tRNA synthetase (AlaRS), glycyl-tRNA synthetase (GlyRS), prolyl-tRNA synthetase (ProRS), histidyl-tRNA synthetase (HisRS), aspartyl-tRNA synthetase

(AspRS), asparaginyl-tRNA synthetases (AsnRS), lysyl-tRNA synthetase-II (LysRS-II), phenylalanyl-tRNA Synthetases (PheRS), PylRS and SepRS), motifs 2 (fRxe) and 3 (gxgxgfd/eR) are responsible for this role. Another class II-conserved motif, motif 1 functions in dimerization of enzyme[2]. Besides these class-specific catalytic domains, both classes contain a ACB domain related to tRNA aminoacylation, and many additional modules with diverse structure and function such as editing.

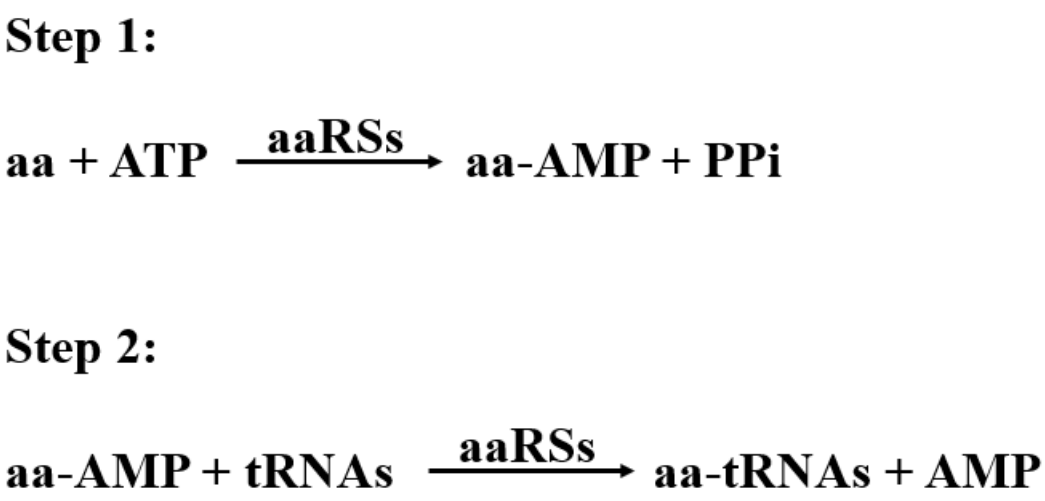


Figure 1.1 Two-steps aminoacylation reaction catalyzed by aminoacyl-tRNA synthetase (AARSs).

1.1.1 The fidelity of amino acid recognition and tRNA recognition

To correctly achieve their tRNA aminoacylation function, aminoacyl-tRNA synthetases (AARSs) need to accurately charge the tRNAs with their cognate amino acids. Thus, it is important for these synthetases to evolve and develop a proper mechanism for the recognition of both substrates. Since the activation of amino acid is the first step of aminoacylation occurring in AARSs, amino acid recognition is one of the most important missions of AARSs. However, it could be challenging for AARSs to recognize and pick up a specific amino acid from a large poor consisting of natural and unnatural amino acids. To overcome this problem and maintain the fidelity of amino acid recognition, AARSs have evolutionally developed multiple strategies, both to secure correct recognition and to avoid the noncognate amino acids[5-7]. Most of the recognition mechanisms rely on the specific non-covalent interaction features in the amino acid binding site of AARSs[1, 2, 8-10]. For instance, the GlyRS uses a highly negatively charged pocket in its active site to avoid the binding of larger amino acids[11], while some AARSs like ThrRS[12] and CysRS[13] use coordinated zinc atom to prevent the recognition of noncognate amino acid. Additionally, by analyzing the features and interaction patterns of the binding site in AARSs, three main points about the amino acid recognition have been summarized as follows: 1) Two classes of AARSs have different overall recognition strategies; 2) Although the binding site and interaction pattern play the most important roles in recognition specificity, very similar amino acids still need more selective strategies like steric effects or editing function; 3) The accurate recognition is indicated as a complicated process which may involves many strategies including the features of binding site, interaction patterns, editing function of enzyme and steric

effects[14]. Furthermore, there are certain AARSs such as GlnRS[15], ArgRS[16] and LysRS[17], requiring cognate tRNA for both recognition and activation of amino acid.

Unlike amino acid, tRNAs usually have very similar base composition and structure to each other. To specifically recognize the cognate tRNA from tRNAs, AARSs not only perform differential binding affinity for their cognate tRNAs (K_m)[18], but also set up a kinetic discrimination to avoid noncognate tRNAs (V_{max})[19]. tRNA identity rules used to explain these effects are also called the second genetic code which relies in each AARS-tRNA pair on the tRNA identity determinants that help cognate tRNA bind to the AARS, or the anti-determinants that prevent false binding of noncognate tRNA to AARS[20-22]. Determinants are mainly located at the two distal ends (anticodon stem or the acceptor arm)[1, 2]. Taking the well-studied alanine system as an example, the G_3-U_{70} identity pair is required for the aminoacylation of tRNA and is highly conserved in different organisms. In addition, the A_{73} discriminator base in acceptor arm is determinant for aminoacyl group transfer[23, 24]. Differently, anti-determinants are not well-studied, and only a few have been reported. One example is the A_{36} in *E. coli* $tRNA^{Arg}$ which help prevent the binding of this tRNA to the TrpRS[25].

1.1.2 Editing function of AARSs and mistranslation

As is mentioned above, AARSs need to distinguish their cognate amino acids from those structurally similar noncognate constitutes. The main amino acid recognition strategies rely more on the feature of the active site and the pattern of interaction between the substrate and active site. However, it is indicated that these strategies are not sufficient for the complete accuracy of cognate amino acids recognition[14, 26, 27]. Indeed, it was estimated that there was an error rate of 10^{-4} in protein synthesis in nature[28]. To face these challenges, AARSs have evolved more complicated mechanisms to ensure accurate amino acid recognition. Editing (proof-reading) function is one of the most useful tools to help part of AARSs ensure the fidelity of amino acid recognition[1, 2, 29]. And, editing function has been found to be more essential for those AARSs which might wrongly activate the structurally similar noncognate amino acids, such as class I IleRS, ValRS, LeuRS and MetRS, as well as class II ThrRS, ProRS, LysRS, AlaRS and PheRS[2, 30]. This proof-reading function allows mischarged noncognate amino acids either to be hydrolyzed from AMP or tRNA through different mechanisms (pre-transfer or post-transfer editing) [1]. The mechanisms for pre-transfer editing are poorly studied but mostly thought to be related to either direct release of mischarged aa-AMPs or hydrolysis of aa-AMPs at active sites or independent editing sites[1, 27]. The editing function found in Class I MetRS[31] is one example for active site hydrolysis, as well as class II LysRS[32]. Post-transfer editing occurs after the transfer of aminoacyl group from the AMP to the tRNA, and it usually involves the translocation of aminoacyl-tRNA from active site to a separated hydrolytic editing domain. Class I IleRS, LeuRS and ValRS perform their post-transfer editing function in CP1 domain, while

class II AARSs tend to have different editing domains and different mechanisms for their post-transfer editing function[2]. ThrRS and AlaRS are thought to deserve more attention because they have structurally related editing sites and they both need to discriminate against the misactivation of serine[33]. Beyond the post-transfer editing, *trans*-editing factor family is also the important participant in editing function. These factors are usually freestanding proteins whose mission is to prevent mischarged tRNAs from protein synthesis in the ribosome. They usually function as extra checkpoints after the amino acids proofreading in AARSs. One of the most well-studied *trans*-editing factors is the AlaX protein which possesses the editing function exclusively against Ser-tRNA^{Ala} [34]. Recently, Vo *et al.* have identified a protein called ANKRD16 in mice which functions like a co-regulator in AlaRS editing process and brings a new editing mechanism that a cofactor can bind directly to the catalytic domain of AlaRS and capture the misactivated serine to prevent the mischarging of serine to tRNA^{Ala}[35].

Although various mechanisms have been developed to help AARSs avoid the potential mischarge of tRNAs by noncognate amino acids, misincorporation of amino acids still exists in organisms from three domains of life. Functional changes occurring at the editing domain can cause amino acid mistranslation. Such an amino acid mistranslation could bring damages to cells, and lead to diseases in mammalian. In mice, it was demonstrated that a missense mutation in AlaRS editing domain could cause mischarge of tRNA^{Ala}, then furtherly triggering the synthesis of misfolded protein and lead to neurodegeneration [36]. In *E. coli*, under severe oxidative stress, ROS like hydrogen peroxide can cause the oxidation of cysteine 182 in the editing domain of ThrRS and impair the editing function of this AARS, which then leads to the

misincorporation of serine by $tRNA^{Thr}$ and damage the growth of cells[37]. In addition, the deficiency of editing function of AARSs can affect the ability of cells to sense the available amino acid pool under amino acids starvation[38]. However, in some instances, amino acid misincorporation can be a beneficial strategy to prevent cell from damage brought by stress conditions. MetRS in yeast and mammalian cells is one of these examples. The posttranslational phosphorylation of MetRS increases under the oxidative stress, resulting in an increase of mismethionylate tRNAs. The excessive amount of Met residues can attract the ROS and then protect the amino acid residues at active site[39]. Similarly, global misacylation of tRNA under the antibiotic stress can help *E. coli* increase their stress resistance[40]. This evidence indicates that mistranslation can be regulated as a tool for certain organisms to deal with various stress conditions.

1.1.3 Noncanonical functions of AARSs

Despite their crucial roles in translation process, AARSs are also found to be involved in a variety of biological processes ranging from gene transcription, RNA splicing to translational regulation[1, 41, 42]. Here, a specific emphasis should be given to ThrRS. In *E. coli*, besides its canonical function in translation, ThrRS also participates in its own mRNA translational regulation. ThrRS can bind to an operator which mimics the anticodon loop of $tRNA^{Thr}$ in the upstream of the initiation codon, leading to the repression of its own mRNA translation[43]. In human cells, ThrRS (or TRS) is recently found to be involved in translation initiation of mRNAs for proteins required for vertebrate development, such as vascular endothelial growth factor

(*VEGF*)[44]. A schematic model for TRS-mediated translation machinery has been proposed. The TRS specifically bind to 4EHP (an eIF4E (the cap-binding protein) homologous protein) via its UNE-T region (eukaryote-unique N-terminal extension) and recruit the initiation factor eIF4A to form an eIF4F-like initiation complex[44]. The selected mRNAs in this model are required to have $tRNA^{Thr}$ anticodon-like loops in their 5' UTR[44]. Another example of noncanonical role is TyrRS which has been proved to function against DNA damage under stress conditions like oxidative stress by activating transcription factor E2F1 to upregulate the expression of DNA damage repair genes[45].

1.2 Lysine acetylation

Protein acetylation, one of the major post-translational modifications (PTMs) in both eukaryotes and prokaryotes, plays important role in biological processes. Normally, there are two different forms of protein acetylation, Nt-acetylation, and lysine acetylation[46]. Nt-acetylation mainly occurs in in N^{α}-termini of the nascent polypeptide chains, catalyzed by Nt-acetyltransferases (NATs)[47], while lysine acetylation is only found in ε-amino group of lysine[46]. Lysine acetylation is a reversible and regulatory process in cells, which is mainly determined by two major groups of enzymes, lysine acetyltransferases (KATs) and lysine deacetylases (KDACs). In addition to these two kinds of enzymes, non-enzymic acetylation (Acetyl-phosphate (AcP) as the donor of acetyl group) has also been demonstrated to be critical in bacterial lysine acetylation[48]. In human cells, heavy attention has been paid to histone acetylation which plays an important role in regulating gene transcription[49]. Similarly, lysine

acetylation has also been identified in prokaryotic nucleoid-associated proteins (NAPs) which is histone-like and function in the regulation of DNA flexibility[50]. Besides these gene regulation-related effects, lysine acetylation is also involved in many different biological processes such as mitochondrial metabolism and protein synthesis.

1.2.1 Lysine acetyltransferases (KATs) and lysine deacetylases (KDACs)

KAT family and KDAC family are the main players in lysine acetylation. The enzymic pathway for lysine acetylation is believed to involve KATs and cofactor acetyl-coenzyme A (acetyl-CoA, the donor of acetyl group). Currently, enzymes from KAT family can be divided into three families: the Gcn5-related N-acetyltransferases (GNAT) family, the p300/CREB-binding proteins (p300/CBP), and the MYST family[46]. GNAT family is one of the most important and conserved KAT. In human or mammalian cells, enzymes in GNAT family like HAT1, GCN5, and PCAF are studied to be related to different histone acetylation[51, 52]. Similarly, the p300/CBP family and the MYST family (containing MOZ, YBF2, SAS2, and TIP60) are also mainly functioning in histone acetylation[52]. Most of KATs in human cells serve as subunits of multiprotein complexes. For example, acetyltransferases p300/CBP can acetylate proteins alone or in the PCAF complex, and they can be autoacetylated in certain lysine positions, which is crucial for their function in H3K53 acetylation[53]. Differently, the KATs in bacteria have been rarely studied. The Gcn5-like acetyltransferase Pat/YfiQ has been identified to be mainly responsible for the enzymic acetylation regulation in *E. coli* and *salmonella* [54, 55]. Recently, in addition to YfiQ, another 4 KATs have been found in *E. coli*, including RimI,

YiaC, YjaB, and PhnO[56].

Lysine deacetylases (KDACs) are those enzymes that can remove the acetyl group from acetyl-lysine. In higher eukaryotic cells, four classes of KDACs (class I, II, III and IV) have been identified. Except class III KDACs (also called sirtuin) requiring NAD^+ as co-substrate to achieve their deacetylase function, all other three classes are found to be Zn^{2+} dependent[57, 58]. Among all Zn^{2+} dependent KDACs, KDAC class I, including HDAC1, 2, 3 and 8, is mostly located at nucleus. KDAC class II (HDAC4, 5, 6, 7, 9, 10) shares high homological catalytic domains with class I[59]. Both class I and class II KDACs mainly function in the deacetylation of histone proteins and several transcription factors[46]. HDAC11 is the only member in class IV, and it's primarily located at nucleus, involved in immune responses[60]. The NAD^+ dependent sirtuins (class III, including sirtuins (SIRT) 1-7) are different to other classes. They are found to localize in many different cellular compartments and participate in various biological processes. Specially, SIRT3, SIRT4 and SIRT5 are predominantly located in mitochondria and influence the energy metabolism[61, 62]. Same as KATs, KDACs in bacteria world need more exploration. The SIRT2 like deacetylase, CobB has been identified in *E. coli* and shows a role in regulating the metabolic processes. However, the mechanism of the regulation is poorly understood[63].

1.2.2 Nonenzymatic lysine acetylation (via acetyl-phosphate, AcP)

It is observed that when transferring *E. coli* cells to media which use glucose or acetate as the nutritional source, cells grow with an increase of acetylation level[48, 64]. But this enhanced acetylation level has been proved to be YfiQ (the only known Gcn5-like acetyltransferase in *E. coli*) independent, suggesting unknown the existence of unknown acetyltransferase or even an acetyltransferase-independent mechanism, which can be the possible reason for this acetylation increasing[48]. To find the potential mechanisms, the focus has been moved to the acetate, a byproduct produced during glucose consumption, and its metabolism. After examining the knock-out strains for three enzymes involved in acetate and acetyl-CoA conversion, acetyl-phosphate (AcP) is demonstrated to be responsible for the glucose-induced acetylation, as the phosphotransacetylase (Pta, converting acetyl-CoA to acetate) knock-out strain failed to accumulate acetylation, while the acetate kinase (AckA, converting acetate to acetyl-CoA) knock-out strain show an increasing of acetylation[48]. In addition, CobB, the deacetylase in *E. coli* shows no preference toward YfiQ dependent acetylation and AcP-mediated acetylation. Furthermore, AcP-mediated acetylation is also found in other bacteria such as *Bacillus subtilis* [65]. As for eukaryotic cells, although nonenzymatic acetylation is not prominent, acetyl-CoA has been indicated to acetylate proteins in mitochondria in a nonenzymatic way[66].

1.2.3 Lysine acetylation and aminoacyl-tRNA synthetases (AARSs)

Previous acetylome studies have shown an enrichment of lysine acetylation in eukaryotic histone proteins, as well as in metabolic pathways such as tricarboxylic acid (TCA) cycle in both eukaryotic and prokaryotic cells[67-69]. It is also remarkable that aminoacyl-tRNA synthetases family is one of the protein family which have the most acetylation site and the highest acetylated level[70-72]. Here, this part will emphasize on the studies related to AARSs acetylation. Recently, several studies have been conducted to learn the effects of lysine acetylation on AARSs with a concentration on their canonical functions. Since K73 residue in motif II of class II AlaRS in *E. coli* is naturally acetylated, a site-specific study has found that the acetylation of K73 can significantly reduce the alanylation activities of the enzyme[73]. Similarly, in class I AARSs, LeuRS and ArgRS, acetylation occurring in catalytic domain (KMSKS motif, K619 of LeuRS, K16 of ArgRS) can also significantly impair the catalytic activity of enzyme. In addition to catalytic domain, acetylated lysine residue in tRNA binding domain can also influence the affinity of LeuRS and ArgRS for tRNA[74]. Moreover, site-specific lysine acetylation (K235 and K238) in KMSKS motif of another class I AARS, TyrRS, is also found to have important effects on the canonical aminoacylation function of enzyme[75]. Furthermore, except MetRS, acetylation is found in KMSKS motif of other class I AARSs[75]. Based on the results from LeuRS, ArgRS and TyrRS, it's expected that acetylation in KMSKS motif can help regulate the canonical activities of class I AARSs. However, more studies are still needed to identify the effects of acetylation on class I and even class II AARSs in *E. coli*. In eukaryotic cells, the effects of acetylation on AARSs are even rarely studied. As is mentioned

above, TyrRS participates in stress response by functioning in transcriptional regulation of DNA damage repair genes[45], however, no possible mechanism to explain the translocation of TyrRS to nuclear has been determined until 2017. Yu's group has reported a mechanism that high acetylation level promotes the translocation of TyrRS[76]. To be specific, an acetyltransferase PCAF (GNAT family) and a class III KDAC, sirtuin 1, are the main players in this mechanism. PCAF can acetylate the K244 residue near nuclear localization signal (NLS) and cause a conformational change, which then promote the localization of TyrRS to nuclear[76]. But this is just an example for the effects of acetylation on noncanonical function of AARSs in eukaryotes, more future investigation will be necessarily needed.

1.3 Genetic code expansion strategy

Naturally, the genetic code of life contains 64 triplet codons, including 3 stop or terminal codons and 61 codons encoding 20 canonical amino acids. Beyond these building basic blocks, natural proteins still require extra various modifications such as posttranslational modifications (PTMs) to function well in organisms. However, it can be challenging to study those modifications, because they are usually dynamic, transient and hard to control[77]. Therefore, to resolve these difficulties, more rigorous and effective strategies to synthesize protein with ideal modifications for different purposes are needed. Rather modify protein post-translationally, co-translationally incorporating noncanonical amino acids (ncAA) into defined sites of protein obviously is a better strategy to specifically modify protein with certain chemical modifications. The first example to incorporate ncAA (selenomethionine (Se-Met)) into proteins in bacteria

during translation dates back to 1956[78]. Since then, lots of ncAAs which can be substrate for natural translational machinery have been successfully incorporated into protein by substituting their natural counterparts[79, 80]. Besides this residue-specific incorporation way, several strategies have been developed to achieve site-specific ncAAs incorporation[81], among which genetic code expansion strategy (GCE) is one the most useful way.

Basically, a pair of engineered orthogonal aminoacyl-tRNA synthetase (AARS) and its orthogonal tRNA, usually from different life domain and not cross-reacting with endogenous AARS/tRNA pair, will be introduced into host cells[82]. Typically, the altered-specificity orthogonal synthetase can recognize the specific ncAA, then aminoacylating the orthogonal tRNA. The ncAA-charged tRNA will be delivered by EF-Tu (elongation factor-Tu) to the ribosome and directly incorporate ncAA into the specific position of object protein in response to an unassigned codon (usually an amber stop codon)(Figure 1.3.1)[83, 84]. Currently, the most widely used othorgonal AARS/tRNA pairs are engineered from either pyrrolysyl-tRNA synthetase (PylRS)/$tRNA^{pyl}$ pair from *Methanosarcina* species[85-87], or tyrosyl-tRNA synthetase from *Methanococcus jannaschii*[88]. Additionally, the codon for original canonical amino acids can be mutant to unassigned codon through site-direct mutagenesis. Thus, genetic code expansion strategy is theoretically applicable to incorporate ncAAs into proteins at any specific positions, which indeed can facilitate the study on the site-specific effects of specific modifications like PTMs on proteins both *in vivo* and *in vitro*.

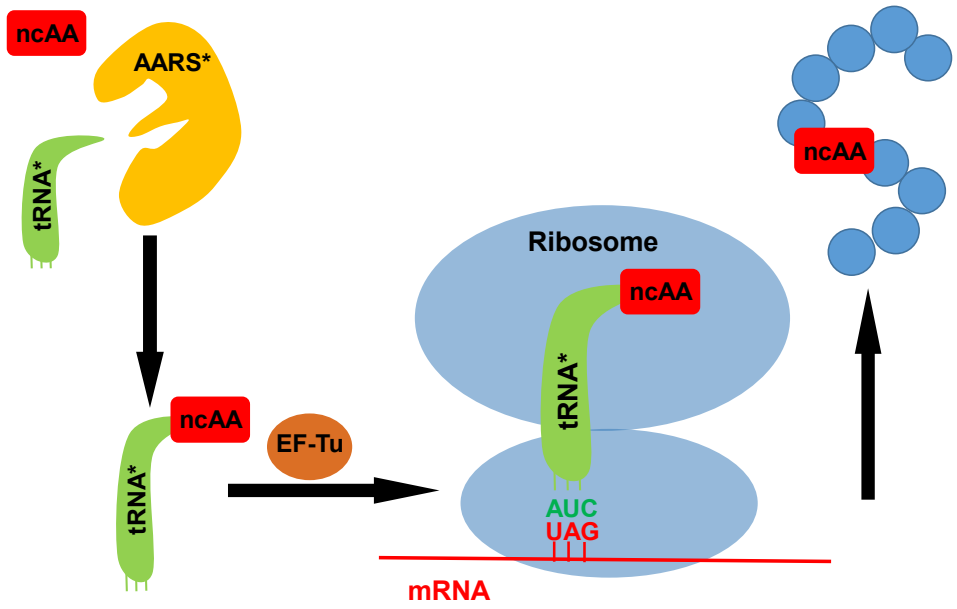


Figure 1.2 A scheme for genetic code expansion strategy[84]. The introduced orthogonal AARS aminoacylates its cognate tRNA with one ncAA. Then the ncAA-charged tRNA is brought to the ribosome by EF-Tu. The introduced tRNA with a designed anticodon can read the corresponding codon in the mRNA (UAG as a representative), then direct the incorporation of ncAA into the specific site of the target protein. AARS*: introduced aminoacyl-tRNA synthetase; tRNA*: introduced tRNA; ncAA: noncanonical amino acid.

1.3.1 Players in GCE strategy

As is mentioned above, there usually are two important factors required in GCE system. One is orthogonal AARS/tRNA pair, and another is the unassigned codon[89, 90]. The first orthogonal AARS/tRNA pair applied to the genetic code expansion of *E.coli* was derived from MjTyrRS-tRNA$_{CUA}$ [91, 92], and this pair was also one of the most frequently used pair in prokaryotic cells. Based on MjTyrRS-tRNA$_{CUA}$, He *et al.* designed three MjTyrRS-opt variants (pBpaRS, pAzPheRS, pIPheRS, respectively) to construct UAA incorporation systems to site specifically incorporate p-benzoyl-L-phenylalanine (pBpa), p-azido-L-phenylalanine (pAzPhe),

and p-iodo-L-phenylalanine (pIPhe) into proteins in *Streptomyces venezuelae*, which would facilitate the study of actinobacterial proteins[93]. Different to MjTyrRS-$tRNA_{CUA}$, PylRS-$tRNA_{CUA}$ pair was used to incorporate pyrrolysine (Pyl), the 22nd proteinogenic amino acid[94] into protein, and it was orthogonal in *E. coli*, yeast, and animals[87, 95-97]. Recently, to make PylRS-$tRNA_{CUA}$ pair system more effective, Fan *et al.* rationally evolved $tRNA^{pyl}$ to create $tRNA^{Pyl}$-opt by setting up three mutant libraries, including library I for the 2-71 and 3-70 positions (replacing these two base pairs with standard base pairs (GC/CG and AU/UA) as well as one wobble pair (GU/UG)), library II for 6-67 and 7-66 with same replacement and library III for 49-65 and 50-64 with same replacement, and the results showed that $tRNA^{Pyl}$-opt could be more effective to incorporate N^{ε}-acetyl-L-lysine into super-folder green fluorescence protein (sfGFP) and implied it as an excellent replacement of wild-type $tRNA^{Pyl}$. Except MjTyrRS-$tRNA_{CUA}$ pair and PylRS-$tRNA_{CUA}$ pair, scientists have also found many other pairs. For example, Park *et al.* site-specifically incorporated O-Phosphoserine (Sep) into a protein in *E.coli*, by designing an Sep-tRNA synthetase (SepRS)-$tRNA^{sep}$ pair and an engineered EF-Tu (EF-Sep)[98]. As for unassigned codons choices, the stop codons and quadruplet codons[99, 100] were often selected. Among them, the amber stop codon UAG was commonly used to direct a noncanonical amino acid at specific position of an aim gene[101].

1.3.2 Recent developments of GCE strategy in lysine acetylation studies

GCE strategy has been used to incorporate a variety of amino acids with specific PTMs (Table 1.3.2), here we give a specific review on studies which develop the GCE system for introducing site-specific lysine acetylation into proteins both in prokaryotic cells and eukaryotic cells.

Recently, the genetic code expansion method for incorporating N^{ε}-acetyllysine (Ack) and its analogs has been developed in bacterial cells. In 2008, Neumann *et al.* demonstrated the site-specific incorporation of Ack in recombinant proteins produced in Escherichia coli using a pyrrolysyl-tRNA synthetase (MbPylRS)/$MbtRNA_{CUA}$ pair and suggested the potential of this strategy in future[102]. Then Yokoyama group successfully introduced this Ack-incorporation system into mammalian cells[103], while the Söll group engineered this system to obtain a AcK-specific PylRS variant (AckRS)[104]. However, these PylRS variants showed serious catalytic deficiency. To solve this problem and optimize the Ack-incorporation system, certain developments have been conducted. Liu group firstly set up a convenient approach which enhanced amber suppression by overexpression of the C-terminal domain of the ribosomal protein L11, to incorporate three three N^{ε}-acetyl-L-lysines (AcKs) into one GFPUV protein in *E. coli*[105]. Besides, both $tRNA^{pyl}$ [106] and PylRS[107] engineering have been performed to achieve high incorporation efficiency. Furthermore, combing optimized $tRNA^{Pyl}$ mutants and PylRS variants, our group develop a facile protocol for Ack incorporation[108].

Table 1.1 Noncanonical amino acids incorporation used in different PTMs studies.

PTM type	Noncanonical Amino Acids
Lysine acetylation	
	Nε-acetyllysine （Ack）
	N^{ε}-thioacetyl-L-lysine (TAcK)
	2-amino-8-oxonoanoica acid (KetoK)
	Trifluoro-acetyllysine (TFAcK)
Lysine ubiquitination	
	D-Cys-ε-Lys
	L-Cys-ε-Lys
	δ-thiol-L-lysine
	δ-hydroxy-L-lysine
Lysine methylation	
	N^{ε}-Boc-N^{ε}-methyllysine
	N^{α}-Fmoc-N^{ε}-(Boc, methyl)-lysine
	N^{ε}-(4-azidobenzoxycarbonyl)-δ, ε-dehydrolysine
	phosphoserine (pSer)
Arginine methylation	
	N^{G}-monomethyl-L-arginine
Tyrosine phosphorylation	
	p-Carboxymethyl-L-phenylalanine (pCMF)
	4-phosphonomethyl-L-phenylalanine (Pmp)
	Phosphotyrosine (pTyr)
	a charge neutral and stable pTyr analog (Uaa 1)
Serine phosphorylation	
	phosphoserine (pSer)
Threonine phosphorylation	
	phosphothreonine (pThr)
Tyrosine sulfation	
	sulfotyrosine
Tyrosine nitration	
	3-nitro-tyrosine (nitro Tyr)
other lysine PTMs	
	ε-N-crotonyllysine (Kcr)
	ε-N-propionyllysine (Kpr)
	ε-N-butyryllysine (Kbu)
	ε-N-alkynyllysine (Kal)
	ε-N-scuccinyllysine (Ksc)
	ε-N-2-hydroxyisobutyryl-lysine (HibK)
Other PTMs	
	3,4-dihydroxy-L-phenylalanine (L-DOPA)

As is mentioned in lysine acetylation part, deacetylases could remove the acetyl group from lysine. Although deacetylase inhibitors like nicotinamine could be helpful, sometimes deacetylase still might affect the Ack incorporation. To avoid the influence from deacetylase, several non-deaceylatable analogs Ack (Figure 1.3.3) were used in certain studies. 2-amino-8-oxononanoic acid (KetoK) was the first Ack analog genetically incorporated into proteins[109]. later, under a cell-free translation condition, thio-acetyllysine (TAcK) was site-specifically incorporated into histone H3 by flexizyme system[110]. Recently, our group engineered the AckRS for better TAcK recognition and also indicated that TAcK could be an ideal mimic of AcK with the ability to avoid deacetylation[111]. In addition to KetoK and TAcK, trifluoro-acetyllysine (TFAcK) was incorporated into p53 protein as a NMR probe for lysine acetylation[112].

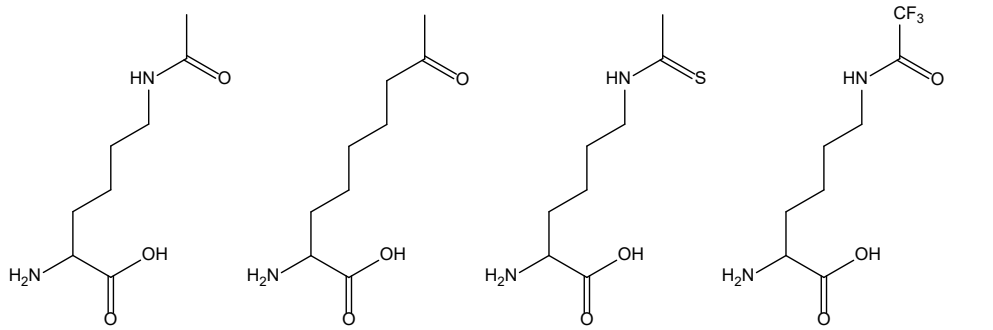


Figure 1.3 The Structures of acetyllysine and its non-deacetylatable analogs[84].

With the development of GCE strategy, researchers were able to study the lysine acetylation of bacteria. For instance, our group was able to use the GCE strategy to study the site-specific effects of lysine acetylation on three important enzymes involved in *E. coli* tricarboxylic acid (TCA) cycle, Malate dehydrogenase (MDH), isocitrate dehydrogenase (ICDH) and type II citrate synthetase (CS)[113-115]. Interestingly, acetylation was able to increase the enzyme activity of MDH and type II CS, whereas it could decrease the enzyme activity of ICDH. Additionally, our group and other lab also applied Ack incorporation system to learn the site-specific effects of lysine acetylation on AARSs in *E. coli*, including class I TyrRS, LeuRS and ArgRS as well as class II AlaRS, and the results showed the acetylation in catalytic domain or tRNA binding domain could affect the catalytic activity of AARSs[73-75]. Similarly, many efforts have been already made to develop the genetic code expansion strategy of eukaryotic cells. In the past, genetic code expansion had been used to incorporate acetyl-lysine into core histones including H2A, H2B and H3[116], however it was hard to get recombinant H4 with acetylation. According to this situation, Wilkins *et al.* improved the general way of acetyl-lysine incorporation by constructing a gene fusion coding for H3 connected to H4 by a proper linker, which would facilitate the production of H4 as same as H3[117]. Except histones, there also were many studies about protein acetylation of other human genes. For instance, Pan *et al.* used the genetic code expansion strategy to incorporate N^{ε}-acetyl-L-lysine into hPrx1at 27 position and found that in different environmental changes, the acetylation in specific site of hPrx1 may change its biological role[118]. As another example, site-specifically K104-acetylated and K107-acetylated K-Ras 4B were synthesized by genetic code expansion method to study K-Ras 4B regulation by

lysine-acetylation[119]. Additionally, Scientists also provided many powerful tool for studying protein acetylation in eukaryotic cells, like mouse[120].

1.4 Rationale and gap in knowledge

Aminoacyl-tRNA synthetases (AARSs) charge tRNAs with cognate amino acids to maintain the faithful process of translation. Some of the AARSs can perform editing or proof-reading function to assure the accuracy of amino acids recognition. Besides, these synthetases also possess noncanonical functions related to several biological processes, such as transcriptional and translational regulation. Recently, proteomic studies have shown an enrichment of lysine acetylation in AARSs. Although a few examples have indicated that acetylation had profound impacts on catalytic activities of AARSs, more studies are needed to well-understand the effects of lysine acetylation on these enzymes. To furtherly understand the relationship between lysine acetylation and AARSs, it's necessary to conduct research on those synthetases whose acetylation haven't been explored before, and it is also important to deeply investigate the potential roles of acetylation in enzyme activities or even other functions, especially the editing function, of AARSs. This study will provide biochemical evidence and structural insight of site-specific effects on AARSs and firstly explore the potential mistranslation caused by acetylation-related editing domain dysfunction of ThrRS.

1.5 References

1. Gomez, M.A.R. and M. Ibba, *Aminoacyl-tRNA synthetases.* RNA, 2020. **26**(8): p. 910-936.

2. Giegé, R. and M. Springer, *Aminoacyl-tRNA synthetases in the bacterial world.* EcoSal Plus, 2016. **7**(1).

3. Mukai, T., et al., *Rewriting the genetic code.* Annual review of microbiology, 2017. **71**: p. 557-577.

4. Cusack, S., *Aminoacyl-tRNA synthetases.* Current opinion in structural biology, 1997. **7**(6): p. 881-889.

5. Mascarenhas, A.P., et al., *Fidelity mechanisms of the aminoacyl-tRNA synthetases*, in *Protein Engineering*. 2009, Springer. p. 155-203.

6. Reynolds, N.M., et al., *Cell-specific differences in the requirements for translation quality control.* Proceedings of the National Academy of Sciences, 2010. **107**(9): p. 4063-4068.

7. Bullwinkle, T., B. Lazazzera, and M. Ibba, *Quality control and infiltration of translation by amino acids outside of the genetic code.* Annual review of genetics, 2014. **48**: p. 149-166.

8. Ibba, M. and D. Söll, *Aminoacyl-tRNA synthesis.* Annual review of biochemistry, 2000. **69**(1): p. 617-650.

9. Ling, J., N. Reynolds, and M. Ibba, *Aminoacyl-tRNA synthesis and translational quality control.* Annual review of microbiology, 2009. **63**: p. 61-78.

10. Hadd, A. and J.J. Perona, *Coevolution of specificity determinants in eukaryotic glutamyl- and glutaminyl-tRNA synthetases.* Journal of molecular biology, 2014. **426**(21): p. 3619-3633.

11. Valencia-Sánchez, M.I., et al., *Structural Insights into the Polyphyletic Origins of Glycyl tRNA Synthetases*♦.* Journal of Biological Chemistry, 2016. **291**(28): p. 14430-14446.

12. Sankaranarayanan, R., et al., *Zinc ion mediated amino acid discrimination by threonyl-tRNA synthetase.* Nature structural biology, 2000. **7**(6): p. 461-465.

13. Zhang, C.-M., et al., *Zinc-mediated amino acid discrimination in cysteinyl-tRNA synthetase.* Journal of molecular biology, 2003. **327**(5): p. 911-917.

14. Kaiser, F., et al., *Characterization of Amino Acid Recognition in Aminoacyl-tRNA Synthetases.* BioRxiv, 2019: p. 606459.

15. Bullock, T.L., et al., *A rationally engineered misacylating aminoacyl-tRNA synthetase.* Proceedings of the National Academy of Sciences, 2008. **105**(21): p. 7428-7433.

16. Delagoutte, B., D. Moras, and J. Cavarelli, *tRNA aminoacylation by arginyl-tRNA synthetase: induced conformations during substrates binding.* The EMBO journal, 2000. **19**(21): p. 5599-5610.

17. Terada, T., et al., *Functional convergence of two lysyl-tRNA synthetases with unrelated topologies.* nature structural biology, 2002. **9**(4): p. 257-262.

18. Lam, S.S.M. and P.R. Schimmel, *Equilibrium measurements of cognate and noncognate interactions between aminoacyl transfer RNA synthetases and transfer RNA.* Biochemistry, 1975. **14**(12): p. 2775-2780.

19. Ebel, J.P., et al., *Factors determining the specificity of the tRNA aminoacylation reaction: Non-absolute specificity of tRNA-aminoacyl-tRNA synthetase recognition and particular importance of the maximal velocity.* Biochimie, 1973. **55**(5): p. 547-557.

20. Pereira, M., et al., *Impact of tRNA modifications and tRNA-modifying enzymes on proteostasis and human disease.* International journal of molecular sciences, 2018. **19**(12): p. 3738.

21. Larkin, D.C., et al., *Identification of essential domains for Escherichia coli tRNALeu aminoacylation and amino acid editing using minimalist RNA molecules.* Nucleic acids research, 2002. **30**(10): p. 2103-2113.

22. Giegé, R., M. Sissler, and C. Florentz, *Universal rules and idiosyncratic features in tRNA identity.* Nucleic acids research, 1998. **26**(22): p. 5017-5035.

23. Musier-Forsyth, K., et al., *Specificity for aminoacylation of an RNA helix: an unpaired, exocyclic amino group in the minor groove.* Science, 1991. **253**(5021): p. 784-786.

24. Fischer, A.E., P.J. Beuning, and K. Musier-Forsyth, *Identification of discriminator base atomic groups that modulate the alanine aminoacylation reaction.* Journal of Biological Chemistry, 1999. **274**(52): p. 37093-37096.

25. Tamura, K., et al., *In vitro study of E. coli tRNAArg and tRNALys identity elements.* Nucleic acids research, 1992. **20**(9): p. 2335-2339.

26. Guo, M. and P. Schimmel, *Structural analyses clarify the complex control of mistranslation by tRNA synthetases.* Current opinion in structural biology, 2012. **22**(1): p. 119-126.

27. Mohler, K. and M. Ibba, *Translational fidelity and mistranslation in the cellular response to stress.* Nature microbiology, 2017. **2**(9): p. 1-9.

28. Kurland, C.G., *Translational accuracy and the fitness of bacteria.* Annual review of genetics, 1992. **26**(1): p. 29-50.

29. Lei, L. and Z.F. Burton, *Evolution of life on Earth: tRNA, aminoacyl-tRNA synthetases and the genetic code.* Life, 2020. **10**(3): p. 21.

30. Jakubowski, H. and E. Goldman, *Editing of errors in selection of amino acids for protein synthesis.* Microbiological reviews, 1992. **56**(3): p. 412-429.

31. Jakubowski, H., *Proofreading in vivo: editing of homocysteine by methionyl-tRNA synthetase in the yeast Saccharomyces cerevisiae.* The EMBO Journal, 1991. **10**(3): p. 593-598.

32. Jakubowski, H., *Misacylation of tRNALys with noncognate amino acids by lysyl-tRNA synthetase.* Biochemistry, 1999. **38**(25): p. 8088-8093.

33. Beebe, K., L.R. De Pouplana, and P. Schimmel, *Elucidation of tRNA-dependent editing by a class II tRNA synthetase and significance for cell viability.* The EMBO Journal, 2003. **22**(3): p. 668-675.

34. Guo, M., et al., *Paradox of mistranslation of serine for alanine caused by AlaRS recognition dilemma.* Nature, 2009. **462**(7274): p. 808-812.

35. Vo, M.-N., et al., *ANKRD16 prevents neuron loss caused by an editing-defective tRNA synthetase.* Nature, 2018. **557**(7706): p. 510-515.

36. Lee, J.W., et al., *Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration.* Nature, 2006. **443**(7107): p. 50-55.

37. Ling, J. and D. Söll, *Severe oxidative stress induces protein mistranslation through impairment of an aminoacyl-tRNA synthetase editing site.* Proceedings of the National Academy of Sciences, 2010. **107**(9): p. 4028-4033.

38. Mohler, K., et al., *Editing of misaminoacylated tRNA controls the sensitivity of amino acid stress responses in Saccharomyces cerevisiae.* Nucleic acids research, 2017. **45**(7): p. 3985-3996.

39. Lee, J.Y., et al., *Promiscuous methionyl-tRNA synthetase mediates adaptive mistranslation to protect cells against oxidative stress.* Journal of cell science, 2014. **127**(19): p. 4234-4245.

40. Schwartz, M.H., et al., *Global tRNA misacylation induced by anaerobiosis and antibiotic exposure broadly increases stress resistance in Escherichia coli.* Nucleic acids research, 2016. **44**(21): p. 10292-10303.

41. Yakobov, N., et al., *Cytosolic aminoacyl-tRNA synthetases: unanticipated relocations for unexpected functions.* Biochimica Et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 2018. **1861**(4): p. 387-400.

42. Mirande, M., *The aminoacyl-tRNA synthetase complex.* Macromolecular Protein Complexes, 2017: p. 505-522.

43. Torres-Larios, A., et al., *Structural basis of translational control by Escherichia coli threonyl tRNA synthetase.* nature structural biology, 2002. **9**(5): p. 343-347.

44. Jeong, S.J., et al., *A threonyl-tRNA synthetase-mediated translation initiation machinery.* Nature communications, 2019. **10**(1): p. 1-15.

45. Wei, N., et al., *Oxidative stress diverts tRNA synthetase to nucleus for protection against DNA damage.* Molecular cell, 2014. **56**(2): p. 323-332.

46. Drazic, A., et al., *The world of protein acetylation.* Biochimica Et Biophysica Acta-Proteins and Proteomics, 2016. **1864**(10): p. 1372-1401.

47. Helsens, K., et al., *Bioinformatics analysis of a Saccharomyces cerevisiae N-terminal proteome provides evidence of alternative translation initiation and post-translational N-terminal acetylation.* Journal of proteome research, 2011. **10**(8): p. 3578-3589.

48. Weinert, B.T., et al., *Acetyl-phosphate is a critical determinant of lysine acetylation in E. coli.* Molecular cell, 2013. **51**(2): p. 265-272.

49. Allfrey, V.G., R. Faulkner, and A.E. Mirsky, *Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis.* Proceedings of the National Academy of Sciences, 1964. **51**(5): p. 786-794.

50. Dillon, S.C. and C.J. Dorman, *Bacterial nucleoid-associated proteins, nucleoid structure and gene expression.* Nature Reviews Microbiology, 2010. **8**(3): p. 185.

51. Allis, C.D., et al., *New nomenclature for chromatin-modifying enzymes.* Cell, 2007. **131**(4): p. 633-636.

52. Montgomery, D.C., A.W. Sorum, and J.L. Meier, *Defining the orphan functions of lysine acetyltransferases.* ACS chemical biology, 2015. **10**(1): p. 85-94.

53. Santos-Rosa, H., et al., *Mechanisms of P/CAF auto-acetylation.* Nucleic acids research, 2003. **31**(15): p. 4285-4292.

54. Wang, Q., et al., *Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux.* Science, 2010. **327**(5968): p. 1004-1007.

55. Starai, V.J. and J.C. Escalante-Semerena, *Identification of the protein acetyltransferase (Pat) enzyme that acetylates acetyl-CoA synthetase in Salmonella enterica.* Journal of molecular biology, 2004. **340**(5): p. 1005-1012.

56. Christensen, D.G., et al., *Identification of novel protein lysine acetyltransferases in Escherichia coli.* MBio, 2018. **9**(5): p. e01905-18.

57. Yang, X.-J. and E. Seto, *The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men.* Nature reviews Molecular cell biology, 2008. **9**(3): p. 206-218.

58. Sauve, A.A., et al., *The biochemistry of sirtuins.* Annu. Rev. Biochem., 2006. **75**: p. 435-465.

59. Mihaylova, M.M. and R.J. Shaw, *Metabolic reprogramming by class I and II histone deacetylases.* Trends in Endocrinology & Metabolism, 2013. **24**(1): p. 48-57.

60. Yanginlar, C. and C. Logie, *HDAC11 is a regulator of diverse immune functions.* Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 2018. **1861**(1): p. 54-59.

61. Yoshida, M., et al., *Chemical and structural biology of protein lysine deacetylases.* Proceedings of the Japan Academy Series B-Physical and Biological Sciences, 2017. **93**(5): p. 297-321.

62. Carabetta, V.J. and I.M. Cristea, *Regulation, Function, and Detection of Protein Acetylation in Bacteria.* Journal of Bacteriology, 2017. **199**(16): p. 15.

63. Weinert, B.T., et al., *Accurate quantification of site-specific acetylation stoichiometry reveals the impact of sirtuin deacetylase CobB on the E. coli acetylome.* Molecular & Cellular Proteomics, 2017. **16**(5): p. 759-769.

64. Schilling, B., et al., *Protein acetylation dynamics in response to carbon overflow in E scherichia coli.* Molecular microbiology, 2015. **98**(5): p. 847-863.

65. Kuhn, M.L., et al., *Structural, kinetic and proteomic characterization of acetyl phosphate-dependent bacterial protein acetylation.* PloS one, 2014. **9**(4): p. e94816.

66. Guan, K.-L. and Y. Xiong, *Regulation of intermediary metabolism by protein acetylation.* Trends in biochemical sciences, 2011. **36**(2): p. 108-116.

67. Choudhary, C., et al., *Lysine acetylation targets protein complexes and co-regulates major cellular functions.* Science, 2009. **325**(5942): p. 834-840.

68. Zhao, S., et al., *Regulation of cellular metabolism by protein lysine acetylation.* Science, 2010. **327**(5968): p. 1000-1004.

69. Nakayasu, E.S., et al., *Ancient regulatory role of lysine acetylation in central metabolism.* MBio, 2017. **8**(6): p. e01894-17.

70. Fan, C. and Q. Gan, *Studying Lysine Acetylation of Aminoacyl-tRNA Synthetases in Escherichia coli.* The FASEB Journal, 2019. **33**(S1): p. 630-3.

71. Wang, G., et al., *Systematic analysis of the lysine acetylome reveals diverse functions of lysine acetylation in the oleaginous yeast Yarrowia lipolytica.* AMB Express, 2017. **7**(1): p. 1-15.

72. Zhang, K., et al., *Comprehensive profiling of protein lysine acetylation in Escherichia coli.* Journal of proteome research, 2013. **12**(2): p. 844-851.

73. Umehara, T., et al., *Lysine acetylation regulates alanyl-tRNA synthetase activity in Escherichia coli.* Genes, 2018. **9**(10): p. 473.

74. Ye, Q., et al., *Acetylation of lysine ϵ-amino groups regulates aminoacyl-tRNA synthetase activity in Escherichia coli.* Journal of Biological Chemistry, 2017. **292**(25): p. 10709-10722.

75. Venkat, S., et al., *Biochemical Characterization of Lysine Acetylation of Tyrosyl-tRNA Synthetase in Escherichia coli.* Chembiochem, 2017.

76. Cao, X., et al., *Acetylation promotes TyrRS nuclear translocation to prevent oxidative damage.* Proceedings of the National Academy of Sciences, 2017. **114**(4): p. 687-692.

77. Lothrop, A.P., M.P. Torres, and S.M. Fuchs, *Deciphering post-translational modification codes.* FEBS letters, 2013. **587**(8): p. 1247-1257.

78. Cohen, G.N. and R. Munier, *Incorporation of structural analogues of amino acids in bacterial proteins.* Biochimica et biophysica acta, 1956. **21**(3): p. 592.

79. Worst, E.G., et al., *Residue-specific Incorporation of Noncanonical Amino Acids into Model Proteins Using an Escherichia coli Cell-free Transcription-translation System.* Journal of Visualized Experiments : JoVE, 2016(114): p. 54273.

80. Johnson, J.A., et al., *Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications.* Current opinion in chemical biology, 2010. **14**(6): p. 774-780.

81. Davis, L. and J.W. Chin, *Designer proteins: applications of genetic code expansion in cell biology.* Nature reviews Molecular cell biology, 2012. **13**(3): p. 168.

82. Ngo, J.T. and D.A. Tirrell, *Noncanonical amino acids in the interrogation of cellular protein synthesis.* Accounts of chemical research, 2011. **44**(9): p. 677-685.

83. Xie, J. and P.G. Schultz, *A chemical toolkit for proteins—an expanded genetic code.* Nature Reviews Molecular Cell Biology, 2006. **7**(10): p. 775-782.

84. Chen, H., et al., *Recent Development of Genetic Code Expansion for Posttranslational Modification Studies.* Molecules, 2018. **23**(7): p. 1662.

85. Chemla, Y., et al., *Genetically expanded cell-free protein synthesis using endogenous pyrrolysyl orthogonal translation system.* Biotechnology and Bioengineering, 2015. **112**(8): p. 1663-1672.

86. Crnković, A., et al., *Pyrrolysyl-tRNA synthetase, an aminoacyl-tRNA synthetase for genetic code expansion.* Croatica Chemica Acta, 2016. **89**(2): p. 163-174.

87. Hancock, S.M., et al., *Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair.* Journal of the American Chemical Society, 2010. **132**(42): p. 14819-14824.

88. Wang, L., et al., *Expanding the genetic code of Escherichia coli.* Science, 2001. **292**(5516): p. 498-500.

89. Zhang, M.S., et al., *Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing.* Nature methods, 2017. **14**(7): p. 729.

90. Terasaka, N., et al., *Recent developments of engineered translational machineries for the incorporation of non-canonical amino acids into polypeptides.* International journal of molecular sciences, 2015. **16**(3): p. 6513-6531.

91. Xie, J. and P.G. Schultz, *An expanding genetic code.* Methods, 2005. **36**(3): p. 227-238.

92. Wang, L., et al., *A New Functional Suppressor tRNA/Aminoacyl− tRNA Synthetase Pair for the in vivo Incorporation of Unnatural Amino Acids into Proteins.* Journal of the American Chemical Society, 2000. **122**(20): p. 5010-5011.

93. He, J., et al., *Development of an unnatural amino acid incorporation system in the actinobacterial natural product producer Streptomyces venezuelae ATCC 15439.* ACS synthetic biology, 2015. **5**(2): p. 125-132.

94. Wan, W., J.M. Tharp, and W.R. Liu, *Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2014. **1844**(6): p. 1059-1070.

95. Namy, O., et al., *Adding pyrrolysine to the Escherichia coli genetic code.* FEBS letters, 2007. **581**(27): p. 5282-5288.

96. Chin, J.W., *Expanding and reprogramming the genetic code of cells and animals.* Annual review of biochemistry, 2014. **83**: p. 379-408.

97. Greiss, S. and J.W. Chin, *Expanding the genetic code of an animal.* Journal of the American Chemical Society, 2011. **133**(36): p. 14196-14199.

98. Park, H.-S., et al., *Expanding the genetic code of Escherichia coli with phosphoserine.* Science, 2011. **333**(6046): p. 1151-1154.

99. O'Donoghue, P., et al., *Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNAPyl for genetic code expansion.* FEBS letters, 2012. **586**(21): p. 3931-3937.

100. Niu, W., P.G. Schultz, and J. Guo, *An expanded genetic code in mammalian cells with a functional quadruplet codon.* ACS chemical biology, 2013. **8**(7): p. 1640-1645.

101. Chin, J.W., *Expanding and reprogramming the genetic code.* Nature, 2017. **550**(7674): p. 53-60.

102. Neumann, H., S.Y. Peak-Chew, and J.W. Chin, *Genetically encoding Nε-acetyllysine in recombinant proteins.* Nature chemical biology, 2008. **4**(4): p. 232-234.

103. Mukai, T., et al., *Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases.* Biochemical and biophysical research communications, 2008. **371**(4): p. 818-822.

104. Umehara, T., et al., *N-acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess N-acetyl lysine specificity in vitro and in vivo.* FEBS letters, 2012. **586**(6): p. 729-733.

105. Huang, Y., et al., *A convenient method for genetic incorporation of multiple noncanonical amino acids into one protein in Escherichia coli.* Molecular BioSystems, 2010. **6**(4): p. 683-686.

106. Fan, C., et al., *Rationally evolving tRNAPyl for efficient incorporation of noncanonical amino acids.* Nucleic acids research, 2015. **43**(22): p. e156-e156.

107. Bryson, D.I., et al., *Continuous directed evolution of aminoacyl-tRNA synthetases.* Nature chemical biology, 2017. **13**(12): p. 1253-1260.

108. Venkat, S., et al., *A Facile Protocol to Generate Site-Specifically Acetylated Proteins in Escherichia Coli.* Journal of visualized experiments: JoVE, 2017(130).

109. Huang, Y., et al., *Genetic incorporation of an aliphatic keto-containing amino acid into proteins for their site-specific modifications.* Bioorganic & medicinal chemistry letters, 2010. **20**(3): p. 878-880.

110. Xiong, H., et al., *Dual genetic encoding of acetyl-lysine and non-deacetylatable thioacetyl-lysine mediated by flexizyme.* Angewandte Chemie International Edition, 2016. **55**(12): p. 4083-4086.

111. Venkat, S., et al., *Genetically encoding thioacetyl-lysine as a non-deacetylatable analog of lysine acetylation in Escherichia coli.* FEBS open bio, 2017. **7**(11): p. 1805-1814.

112. Zhang, F., et al., *A genetically encoded 19 F NMR probe for lysine acetylation.* Chemical Communications, 2018. **54**(31): p. 3879-3882.

113. Venkat, S., et al., *Studying the lysine acetylation of malate dehydrogenase.* Journal of molecular biology, 2017. **429**(9): p. 1396-1405.

114. Venkat, S., et al., *Characterizing lysine acetylation of isocitrate dehydrogenase in Escherichia coli.* Journal of molecular biology, 2018. **430**(13): p. 1901-1911.

115. Venkat, S., et al., *Characterizing lysine acetylation of Escherichia coli type II citrate synthase.* The FEBS journal, 2019.

116. Neumann, H., et al., *A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation.* Mol Cell, 2009. **36**(1): p. 153-63.

117. Wilkins, B.J., et al., *Genetically encoding lysine modifications on histone H4.* ACS chemical biology, 2015. **10**(4): p. 939-944.

118. Pan, Y., et al., *Significant enhancement of hPrx1 chaperone activity through lysine acetylation.* Chembiochem, 2014. **15**(12): p. 1773-1776.

119. Knyphausen, P., et al., *Insights into K-Ras 4B regulation by post-translational lysine acetylation.* Biological chemistry, 2016. **397**(10): p. 1071-1085.

120. Han, S., et al., *Expanding the genetic code of Mus musculus.* Nature communications, 2017. **8**: p. 14568.

CHAPTER II: Site-specifically Studying Lysine Acetylation of Aminoacyl-tRNA Synthetases

Hao Chen[1], Sumana Venkat[1], Denver Hudson[2], Tony Wang[3], Qinglei Gan[2], Chenguang Fan[1,2]*

[1]Cell and Molecular Biology Program, [2]Department of Chemistry and Biochemistry,

[3]Department of Biological Sciences, University of Arkansas, Fayetteville, AR 72701, USA

2.1 Abstract

Aminoacyl-tRNA synthetases (AARSs) charge their cognate tRNAs with corresponding amino acids, playing key roles in ribosomal protein synthesis. A series of proteomic studies have demonstrated that AARSs have much higher levels of lysine acetylation than other proteins in *Escherichia coli*. To study AARS acetylation, 25 site-specifically acetylated variants of four AARSs were generated by the genetic code expansion strategy. Kinetic analyses were performed to biochemically characterize the impact of site-specific acetylation on AARS functions, including amino acid activation, tRNA aminoacylation, and editing activities. The results showed that impacts of acetylation were different between class I and class II AARSs, and also varied among the same class of AARSs. The results also showed that acetylation of threonyl-tRNA synthetase (ThrRS) could affect its editing function. Both *in vivo* and *in vitro* studies were further performed to explore the acetylation and deacetylation processes of ThrRS. Although nonenzymatic acetylation and CobB-dependent deacetylation were concluded, the results also indicated existence of additional modifying enzymes or mechanisms for ThrRS acetylation and deacetylation.

2.2 Introduction

The first step of ribosomal protein synthesis is to charge tRNAs with their cognate amino acids. This reaction is catalyzed by an ancient family of enzymes known as aminoacyl-tRNA synthetases (AARSs), and proceeds with two steps: activating a specific amino acid and transferring it to the correct tRNA [1]. Due to the essential role of translation in all living species, the AARS family is one of the most ubiquitous and conserved protein families in all three domains of life, and their functions are critical for the high fidelity of the translation process [2]. Besides this canonical role, AARSs are also found to be involved in many other biological processes such as transcription, translation, RNA splicing, cell signaling and migration [3, 4], and are associated with a variety of human diseases including mitochondrial diseases, Charcot-Marie-Tooth and related neuropathies, cancers, and infectious diseases [5-7], making AARSs favorable targets for diagnostic and therapeutic development [8-10].

For both basic sciences and medical applications, a number of studies have been performed to explore AARS regulation, most of which focused on gene expression of AARSs [11]. Different from regulation at the gene expression level which needs relatively longer time, post-translational modifications (PTMs) can rapidly modify properties of existing proteins to regulate their functions for adapting to different stimuli or environmental changes [12]. It has been shown that phosphorylation of eukaryotic AARSs mostly affects noncanonical functions of AARSs, and only has little influence on their canonical tRNA charging activities [13, 14]. Recently, we and others showed that acetylation of lysine residues in the KMSKS motifs of three Escherichia coli class I AARSs, tyrosyl-tRNA synthetase (TyrRS), leucyl-tRNA synthetase (LeuRS), and arginyl-

tRNA synthetase (ArgRS), could impair their aminoacylation activities [15, 16]. Although class II AARSs have no KMSKS motifs, acetylation of the lysine residue at the tRNA binding site could also decrease the activity of *E. coli* alanyl-tRNA synthetase (AlaRS) [17]. In human cells, acetylation of TyrRS could promote its nuclear translocation to prevent oxidative damage under oxidative stress [18]. These studies demonstrated the role of acetylation in regulating both canonical and noncanonical functions of AARSs.

To study lysine acetylation, glutamine is commonly used to be a mimic of acetyllysine, which however does not always yield expected results [19]. To overcome this issue, the genetic code expansion strategy has been applied to use an engineered pyrrolysyl-tRNA synthetase variant specific for acetyllysine (AcK) and its cognate $tRNA^{Pyl}$ to incorporate AcK directly at an assigned codon to produce site-specifically acetylated proteins [20]. To explore general mechanisms of AARS acetylation, we utilized the genetic code expansion strategy to generate totally 25 purely acetylated variants of four *E. coli* AARSs, and characterized their site-specific impacts on amino acid activation, tRNA aminoacylation, and editing activities. The acetylation and deacetylation processes were also investigated.

2.3 Methods

2.3.1 Experimental materials

Chemicals and bacterial growth media were purchased from Sigma-Aldrich (St. Louis, MO), BDH Chemicals (Radnor, PA) or CHEM-IMPEX (Wood Dale, IL). Radioactive compounds were purchased from PerkinElmer (Waltham, MA) or VWR International (Radnor,

PA). BL21 (DE3) cells, DH5α cells, Q5 site-directed mutagenesis kits, DNA assembly kits, and DpnI enzymes were purchased from New England Biolabs (Ipswich, MA). The Bradford protein assay, the Bio-Safe Coomassie stain, SDS-PAGE ready gels, Biospin 30 columns, and the Trans-Blot Turbo Transfer System were purchased from Bio-Rad (Hercules, CA). DH5α cells were used for cloning in this work. Plasmids were constructed from PCR fragments by DNA assembly kits. Stop codon mutations were made by Q5 site-directed mutagenesis kits. Strains and plasmids used in this study is listed in Table S1. *E. coli* single-knockout strains were from the Keio collection [21]. Inactivation of additional genes from the *E. coli* genome was performed by the recombination approach [22]. CobB, RcsB, and YfiQ proteins were expressed from the ASKA collection and purified by Ni-NTA affinity chromatography [23]. Expression and purification of AARSs and their acetylated variants as well as SDS-PAGE, western blotting, and MS analyses followed previous protocols (see details in Supporting Information) [24-26].

2.3.2 Enzyme assays

The ATP-PPi exchange assay was modified from previous experiments [15]. A reaction (25 μL) contained the following components: HEPES-KOH (100 mM, pH 7.5), KCl (30 mM), $MgCl_2$ (10 mM), DTT (2 mM), KF (2 mM), [^{32}P]-NaPPi (2 mM), ATP (5 mM or varied from 0.1 - 10 KM for kinetic analyses), individual amino acids (5 mM or varied from 0.1 - 10 KM for kinetic analyses; for ThrRS misactivation, serine was varied from 5 - 500 mM), and AARS or their acetylated variants (100 nM). Reaction mixtures were incubated at 37°C. At each time points, 1 μL of aliquots from reaction mixtures were plotted immediately to PEI-cellulose plates

(Merck, Kenilworth, NJ). Reaction mixtures were separated on PEI-cellulose plates in 1 M urea and 1 M monopotassium phosphate. Plates were then scanned by a phosphorimager. Kinetic parameters were analyzed by non-linear regression with software GraFit.

The aminoacylation assay was modified from previous experiments [15]. *E. coli* tRNAs were generated by modified protocols from previous experiments (see details in Supporting Information) [27]. A reaction mixture (50 μL) containing HEPES-KOH (100 mM, pH 7.2), KCl (30 mM), $MgCl_2$ (10 mM), ATP (5 mM, pH 7.0), DTT (5 mM), individual amino acids (200 μM [^{35}S]-cysteine, 100 μM [^{3}H]-aspartic acid, 50 μM [^{3}H]-histidine, 50 μM [^{14}C]-threonine, or 50 μM [^{3}H]-serine), tRNA (10 μM or varied from 0.1 - 10 KM for kinetic analyses), and AARSs or their variants (10 nM or 5 μM in ThrRS mischarging reactions) was incubated at 37°C. Reaction mixtures (10 μL) were spotted on filter paper pre-soaked with trichloroacetic acid (5%, TCA) at each time points. The paper discs were washed three times with TCA (5%), rinsed with ethanol (95%), dried at 80°C for 20 min, and read by scintillation counting. For each sample, blank reaction (containing no enzyme) was set as background. Kinetic parameters were analyzed by non-linear regression with software GraFit.

In vitro acetylation and deacetylation assays followed previous experiments (see details in Supporting Information) [28].

2.4 Results and discussion

AARSs can be divided into two classes based on mutually exclusive sequence motifs and different active-site structures [29, 30]. In class I AARSs, signature motifs HIGH and KMSKS are responsible for substrate binding and intermediate stabilization [31]. The acetylation of three class I AARSs have been studied by us and others [15, 16], so in this work we only tested one additional class I cysteinyl-tRNA synthetase (CysRS). Different from class I AARS, class II AARSs have no HIGH and KMSKS motifs. Instead, they contain motifs 2 (fRxe) and 3 [gxgxgf(d/e)R] to play the role in ATP binding [31, 32]. Rather than lysine, both motifs contain conserved arginine residues which are also positively charged in physiological conditions but cannot be acetylated. In this study, we tested three class II AARSs including aspartyl-tRNA synthetase (AspRS), histidyl-tRNA synthetase (HisRS), and threonyl-tRNA synthetase (ThrRS). All the four AARSs selected have solved crystal structures for illustrating the impacts of acetylation on AARS functions.

Although a number of *E. coli* acetylome studies have been performed, those identified lysine acetylation targets are not identical, mostly resulting from the differences in *E. coli* strains, growth conditions, sensitivities of acetyllysine antibodies, and peptide-detecting methods [33]. Moreover, according to several quantitative mass spectrometry (MS) analyses, the site-specific stoichiometry of acetylation is different from protein-to-protein and site-to-site [34-36]. Previous AARS acetylation studies chose acetylation sites mostly by their known functions [15-17], which might not reflect the real AARS acetylation *in vivo*. To study acetylation with more biological relevance, we selected those that have the highest acetylation stoichiometry *in vivo* and have

been identified in at least three independent proteomic studies. Totally, we selected 25 acetylation sites in those four AARSs for further studies.

2.4.1 Biochemical characterization of acetylation effects on AARS activities

1) Class I CysRS

For CysRS, K73, K76, K175, K269, and K310 were selected for site-specific acetylation studies. All of them are on the protein surface (Figure S1A). K175 and K269 are located in the CP domain and the KMSKS motif, respectively, both of which are conserved features of class I AARSs [37]. The optimized acetyllysine (AcK) incorporation system was used to generate site-specifically acetylated CysRS variants (Figure S1B) [38]. The purity and incorporation sites were confirmed with full-length MS and LC-MS/MS analyses, respectively (Figure S2-S7). Kinetic analyses were performed on the wild-type CysRS and its acetylated variants to compare their activities of amino acid activation and tRNA charging (Table 2.1).

The acetylation of K73 and K76 had no obvious impacts on both cysteine activation and tRNA charging. The acetylation of K175 had no significant effects on cysteine activation, but decreased the catalytic efficiency of tRNA charging by 5-fold, possibly because K175 is located in a loop region which stabilizes the hairpin of the 3' end UCCA-76 of the $tRNA^{Cys}$. As the conserved lysine residue in the KMSKS motif, acetylation of K269 decreased the catalytic efficiencies of both cysteine activation and tRNA charging by 5-10 folds. Our previous work on TyrRS acetylation showed that the acetylation of lysine residues in the KMSKS motif could impair the activity by more than 200 folds [15]. The relieved decrease in CysRS may result from

the help of the zinc ion which plays the key role in amino acid discrimination [39]. The CD spectrum of the acetylated variant at K269 was similar to that of the wild-type CysRS, indicating that the acetylation of K269 affects cysteine activation and tRNA charging directly rather than the structural stability of the enzyme (Figure S8). Located at the binding region with the $tRNA^{Cys}$ anticodon stem, the acetylation of K310 caused a slight decrease in tRNA binding.

Table 2.1 Steady-state kinetic parameters of *E. coli* CysRS and its acetylated variants*.

	Cysteine			ATP			$tRNA^{Cys}$		
	K_M (μM)	k_{cat} (S^{-1})	k_{cat} / K_M (S^{-1} μM^{-1})	K_M (μM)	k_{cat} (S^{-1})	k_{cat} / K_M (S^{-1} μM^{-1})	K_M (μM)	k_{cat} (S^{-1})	k_{cat} / K_M (S^{-1} μM^{-1})
WT	27 ± 3	88 ± 7	3.3	271 ± 13	95 ± 4	0.35	1.4 ± 0.2	2.9 ± 0.4	2.1
73AcK	35 ± 1	91 ± 6	2.6	263 ± 17	99 ± 8	0.38	1.2 ± 0.1	2.3 ± 0.2	1.9
76AcK	26 ± 5	84 ± 2	3.2	303 ± 13	87 ± 3	0.29	1.3 ± 0.0	2.6 ± 0.5	2.0
175AcK	33 ± 2	78 ± 4	2.4	364 ± 26	81 ± 6	0.22	2.3 ± 0.6	0.9 ± 0.3	0.39
269AcK	55 ± 9	34 ± 3	0.62	872 ± 35	29 ± 4	0.03	1.7 ± 0.2	0.4 ± 0.0	0.24
310AcK	25 ± 3	86 ± 9	3.4	307 ± 18	88 ± 11	0.29	2.2 ± 0.7	2.5 ± 0.4	1.1

* Kinetic parameters for cysteine and ATP were determined by ATP-PPi-exchange assays. Kinetic parameters for $tRNA^{Cys}$ were determined by aminoacylation assays. Mean and standard deviations were calculated from three biological replicates.

2) Class II AspRS

For AspRS, K81, K186, K283, K332, and K353 were selected for site-specific acetylation studies. AspRS forms homodimer, and all selected sites are on protein surface except K283 (Figure S9A and S9B) [40]. The optimized acetyllysine (AcK) incorporation system was used to generate site-specifically acetylated AspRS variants (Figure S9C). The purity and incorporation sites were confirmed with MS analyses (Figure S10-S15). Kinetic analyses were performed on the wild-type AspRS and its acetylated variants (Table 2.2).

The acetylation of K186, K332, and K353 had no obvious effects on both aspartic acid activation and $tRNA^{Asp}$ charging. Although K81 is located at the anticodon recognition region, its acetylation only caused a slight increase in the K_M value of tRNA. Acetylation of K283 also had no significant impacts on AspRS activities even it is close to the active site. These site-specific results indicated that lysine acetylation had no significant impacts on AspRS functions. K198 was previously shown to be essential in aspartic acid activation [41]. However, the acetylation of K198 has not been detected in cells by proteomic studies.

Table 2.2 Steady-state kinetic parameters of *E. coli* AspRS and its acetylated variants*.

	Aspartic acid			ATP			$tRNA^{Asp}$		
	K_M (μM)	k_{cat} (S^{-1})	k_{cat} / K_M (S^{-1} mM^{-1})	K_M (μM)	k_{cat} (S^{-1})	k_{cat} / K_M (S^{-1} mM^{-1})	K_M (μM)	k_{cat} (S^{-1})	k_{cat} / K_M (S^{-1} μM^{-1})
WT	512 ± 21	25 ± 3	49	620 ± 32	31 ± 3	50	0.5 ± 0.1	7.5 ± 0.9	15
81AcK	506 ± 38	24 ± 2	47	657 ± 45	28 ± 1	43	0.8 ± 0.2	7.0 ± 1.2	8.8
186AcK	537 ± 29	27 ± 6	50	612 ± 29	33 ± 2	54	0.5 ± 0.0	7.2 ± 0.3	14
283AcK	513 ± 47	23 ± 4	45	674 ± 21	28 ± 2	42	0.6 ± 0.1	6.5 ± 0.9	11
332AcK	505 ± 31	26 ± 2	51	603 ± 47	29 ± 4	48	0.5 ± 0.1	7.1 ± 0.6	14
353AcK	549 ± 12	27 ± 1	49	638 ± 18	32 ± 5	50	0.5 ± 0.2	6.8 ± 0.4	14

* Kinetic parameters for aspartic acid and ATP were determined by ATP-PPi-exchange assays. Kinetic parameters for $tRNA^{Asp}$ were determined by aminoacylation assays. Mean and standard deviations were calculated from three biological replicates.

3) Class II HisRS

For HisRS, K30, K53, K65, K189, and K378 were selected for site-specific acetylation studies. HisRS also forms homodimer, and K30 and K378 are located at the dimer interface, while K53 and K65 are at protein surface (Figure S16A) [42]. K189 is in the insertion domain which was disordered in the crystal structure. As there is no solved structure of *E. coli* HisRS in complex with $tRNA^{His}$, the structure of such complex from Thermus thermophilus was used to demonstrate the position of $tRNA^{His}$ (Figure S16B) [43]. The optimized acetyllysine (AcK) incorporation system was used to generate site-specifically acetylated HisRS variants (Figure S16C). The purity and incorporation sites were confirmed with MS analyses (Figure S17-S22). Kinetic analyses were performed on the wild-type HisRS and its acetylated variants (Table 2.3).

Although K30 is located at the dimer interface, its acetylation had no obvious effects on histidine activation and $tRNA^{His}$ charging, possibly because it is the hydrophobic interaction that plays the major role in subunit interactions [44]. K53 and K65 are at the entrance of the active site, but their acetylation also had no significant impacts on HisRS activities. K189 is in the insertion domain which binds to the acceptor stem of $tRNA^{His}$, but its acetylation of K189 had no effects on both histidine activation and tRNA aminoacylation. The acetylation of K378 resulted in a 5-fold decrease in the catalytic efficiency of tRNA charging. According to the crystal structure of T. thermophilus HisRS complexed with $tRNA^{His}$, K375 (the counterpart of K378) interacts with the anticodon stem (Figure S16B), so the decreased activity of the 378AcK variant was probably caused by impaired binding with the $tRNA^{His}$. The CD spectrum of the 378AcK variant was similar to that of the wild-type HisRS, indicating that the acetylation of K378 affects

tRNA charging directly rather than the structure of the enzyme (Figure S23).

Table 2.3 Steady-state kinetic parameters of *E. coli* HisRS and its acetylated variants*.

	Histidine			ATP			$tRNA^{His}$		
	K_M (μM)	k_{cat} (S^{-1})	k_{cat} / K_M (S^{-1} μM^{-1})	K_M (μM)	k_{cat} (S^{-1})	k_{cat} / K_M (S^{-1} μM^{-1})	K_M (μM)	k_{cat} (S^{-1})	k_{cat} / K_M (S^{-1} μM^{-1})
WT	32 ± 4	120 ± 23	3.8	763 ± 49	145 ± 37	0.19	0.7 ± 0.2	3.4 ± 0.5	4.9
30AcK	36 ± 2	113 ± 19	3.1	737 ± 57	132 ± 7	0.18	0.7 ± 0.1	3.6 ± 1.1	5.1
53AcK	33 ± 7	109 ± 38	3.3	807 ± 32	115 ± 12	0.14	0.8 ± 0.3	3.1 ± 0.7	3.9
65AcK	33 ± 9	115 ± 29	3.5	864 ± 28	126 ± 24	0.15	0.9 ± 0.2	3.0 ± 0.4	3.3
189AcK	29 ± 1	111 ± 16	3.8	735 ± 42	123 ± 15	0.17	0.7 ± 0.2	3.2 ± 0.3	4.6
378AcK	46 ± 6	107 ± 8	2.2	890 ± 75	109 ± 10	0.12	1.7 ± 0.5	1.5 ± 0.2	0.9

* Kinetic parameters for histidine and ATP were determined by ATP-PPi-exchange assays. Kinetic parameters for $tRNA^{His}$ were determined by aminoacylation assays. Mean and standard deviations were calculated from three biological replicates.

4) Class II ThrRS

For ThrRS, K122, K130, K169, K197, K200, K226, K286, K426, K577, and K599 were selected for site-specific acetylation studies. Except K197 and K200, all other lysine residues are at the protein surface (Figure S24A) [45]. Like AspRS and HisRS, ThrRS also forms homodimer, and K286 is at the subunit interface (Figure S24B). K122, K130, K169, K197, and K200 are located in the editing domain (Figure S24C), while K226 is in the helical linker which connects the editing domain and the catalytic core. The optimized acetyllysine (AcK) incorporation system was used to generate site-specifically acetylated ThrRS variants (Figure S24D). The purity and incorporation sites were confirmed with MS analyses (Figure S25-S35). Kinetic analyses were performed on the wild-type HisRS and its acetylated variants (Table 2.4).

The acetylation of K200 had a 3-fold decrease in the catalytic efficiency of $tRNA^{Thr}$ charging, mostly because of its direct contact with the tRNA [46] (Figure S24A), while acetylation of other lysine residues located at the editing domain and the linker helix had no obvious effects. K286 is in the class II-conserved motif 1 which forms the major part of the dimer interface and interacts with the C-terminal of the other subunit which binds to the tRNA anticodon region [45], and its acetylation had a 2-fold decrease in the catalytic efficiency of tRNA aminoacylation. Although K426 is located at the entrance of the catalytic cleft, its acetylation had no significant impacts on both threonine activation and tRNA aminoacylation. K577 and K599 are located at the anticodon loop binding region. K599 is close to G35, while K577 inserts into the open space of the anticodon loop created by splayed-out bases G35 to A38 [45]. The acetylation of both residues resulted in 2-3 fold of increases in the KM values for tRNA. The CD spectrum of the 577AcK variant was similar to that of the wild-type ThrRS, indicating that the acetylation of K577 affects tRNA charging directly rather than the proper folding of the enzyme (Figure S36).

Table 2.4 Steady-state kinetic parameters of *E. coli* ThrRS and its acetylated variants*.

	Threonine			ATP			$tRNA^{Thr}$		
	K_M (μM)	k_{cat} (S^{-1})	k_{cat} / K_M (S^{-1} μM^{-1})	K_M (mM)	k_{cat} (S^{-1})	k_{cat} / K_M (S^{-1} mM^{-1})	K_M (μM)	k_{cat} (S^{-1})	k_{cat} / K_M (S^{-1} μM^{-1})
WT	102 ± 8	27 ± 3	0.26	342 ± 29	26 ± 4	0.076	0.9 ± 0.4	0.7 ± 0.2	0.78
122AcK	99 ± 6	29 ± 6	0.29	312 ± 47	29 ± 2	0.093	1.0 ± 0.3	0.7 ± 0.1	0.70
130AcK	108 ± 12	26 ± 5	0.24	335 ± 19	26 ± 5	0.078	1.0 ± 0.2	0.7 ± 0.2	0.70
169AcK	97 ± 5	25 ± 5	0.26	309 ± 32	24 ± 3	0.078	0.9 ± 0.3	0.6 ± 0.3	0.66
197AcK	112 ± 18	23 ± 4	0.21	306 ± 42	22 ± 6	0.072	1.2 ± 0.5	0.7 ± 0.3	0.58
200AcK	108 ± 9	27 ± 2	0.25	397 ± 43	25 ± 7	0.063	2.1 ± 0.7	0.5 ± 0.2	0.24
226AcK	105 ± 17	24 ± 3	0.23	338 ± 25	24 ± 6	0.071	1.1 ± 0.1	0.7 ± 0.1	0.63
286AcK	120 ± 21	22 ± 2	0.18	370 ± 17	22 ± 1	0.059	1.2 ± 0.3	0.5 ± 0.0	0.41
426AcK	109 ± 14	25 ± 1	0.23	354 ± 36	24 ± 3	0.067	1.0 ± 0.1	0.6 ± 0.2	0.60
577AcK	96 ± 10	26 ± 3	0.27	321 ± 21	26 ± 4	0.081	3.2 ± 0.6	0.6 ± 0.0	0.19
599AcK	113 ± 27	25 ± 2	0.22	366 ± 30	26 ± 2	0.071	2.3 ± 0.4	0.6 ± 0.1	0.26

* Kinetic parameters for threonine and ATP were determined by ATP-PPi-exchange assays. Kinetic parameters for $tRNA^{Thr}$ were determined by aminoacylation assays. Mean and standard deviations were calculated from three biological replicates.

2.4.2 Characterization of acetylation effects on the ThrRS editing activity

Besides aminoacylation activity, ThrRS also has the editing activity to deacylate mischarged Ser-tRNA^{Thr} [47]. So, we further characterized the effect of lysine acetylation on ThrRS editing activities. First, we tested the serine misactivation by ThrRS. Without any key residues in the amino acid binding site, all acetylated variants had similar serine activation activities with the wild-type ThrRS (Table S2). Then, we measured the generation of mischarged Ser-tRNA^{Thr} (Figure 2.1). K156 is one of the key residues in the editing site of ThrRS [46]. However, the acetylation of K156 has not been identified in cells by proteomic studies. In this work, we used the ThrRS-K156A variant as a comparison. Surprisingly, the ThrRS-169AcK variant could also generate mischarged Ser-tRNA^{Thr}, which was about half of that produced by the ThrRS-K156A variant. According to the crystal structure, K169 is located at the opening of the editing site (Figure S24C), and its acetylation may affect the proper orientation of Ser-tRNA^{Thr} for deacylation.

In previous studies of acetylation in LeuRS and AlaRS, which both have editing domains, but no lysine acetylation was shown to be associated with editing functions [16, 17]. This work is the first to show the association of AARS acetylation with its editing activity. Besides lysine acetylation, oxidation of C182 of ThrRS has been shown to impair the editing activity under oxidative stress [48]. Since lysine acetylation is commonly existing during regular physiology conditions, the impact of acetylation on AARS editing could be more comprehensive. As the acetylation of ThrRS may cause protein mistranslation, its acetylation process was then further investigated.

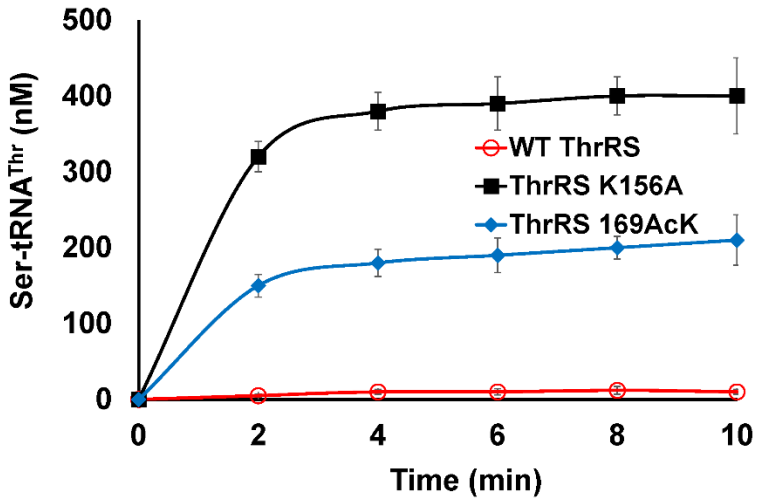


Figure 2.1 Serylation by *E. coli* ThrRS and its variants. 5 μM in ThrRS or its variants, 50 μM [^{3}H]-serine, 10 μM purified tRNA^{Thr}, and 5 mM ATP was in the reaction mixture and incubated at 37°C. Other acetylated variants had no measurable activities, so they were not shown in the figure. Mean and standard deviations were calculated from three biological replicates. Open circle: the wild-type ThrRS; Closed diamond: the ThrRS 169AcK variant; Closed square: the ThrRS K156A variant.

2.4.3 The acetylation process of *E. coli* ThrRS

Lysine acetyltransferases (KATs) catalyze the reaction of transferring an acetyl group from acetyl-CoA (AcCoA) to the ε-amino group of a lysine residue in proteins, including three major families: Gcn5-related N-acetyltransferase (GNAT) family, the p300/CBP family, and the MOZ, YBF2, SAS2 and TIP60 (MYST) family [49]. The p300/CBP and MYST families are present only in eukaryotes, while the GNAT family is distributed across all three domains of life. *E. coli* has 25 GNAT homologues, but YfiQ is the only well-studied KAT [50]. So, we first tested the acetylation activity of YfiQ on ThrRS both *in vivo* and *in vitro*.

For *in vivo* studies, the native ThrRS was purified from the BW25113 wild-type or Δ*yfiQ* cells grown in M9 media with 0.2% glucose by a modified protocol [51]. Their acetylation levels were compared by western blotting (Figure S37A). The weak band for the ThrRS from the wild-type cells indicated a low acetylation level. The deletion of *yfiQ* did not decrease the acetylation level of ThrRS expressed in wild-type cells, indicating that ThrRS is not the substrate of YfiQ *in vivo*. For *in vitro* studies, the wild-type ThrRS with a His_6-tag was expressed and purified by affinity chromatography from BL21 (DE3) cells, then treated with purified YifQ and AcCoA. After 1-hour incubation, ThrRS had no detectable acetylation, indicating that ThrRS is not the substrate of YfiQ *in vitro* either (Figure S37B). Then we further tested the acetylation of ThrRS by the other 24 GNAT family members, and no obvious acetylation was detected (Figure S38).

Besides AcCoA-dependent enzymatic acetylation, a number of studies have proposed an enzyme-independent mechanism which uses AcCoA or acetyl-phosphate (AcP) as the acetyl group donor to acetylate lysine residues chemically [52]. Proteomic studies even showed that

AcP-dependent acetylation overwhelms YfiQ-dependent acetylation in *E. coli* [53, 54]. Thus we tested chemical acetylation of ThrRS by AcP *in vitro*. The wild-type ThrRS expressed and purified from BL21 (DE3) cells was treated with AcP at concentrations of 200 μM, 3 mM, and 12 mM, which correspond to intracellular AcP concentrations at the exponential phase, the stationary phase, and the AcP-accumulating Δ*ackA* strain, separately (Those concentration is not unusually high, as phosphotransacetylase (Pta) and acetate kinase (AckA) can generate AcP at concentrations around 50-60 mM. [55]) The acetylation levels of AcP-treated ThrRS were determined by western blotting (Figure 2A). Results showed that AcP could chemically acetylate ThrRS in dose- and time-dependent manners *in vitro*. Next, we performed *in vivo* tests for AcP-dependent acetylation of ThrRS. AcP is the intermediate of the Pta-AckA pathway [56]. Pta produces AcP from AcCoA, and AckA converts AcP to acetate. So *Δpta* blocks the AcP production, while *ΔackA* accumulates AcP. The acetylation levels of native ThrRS purified from BW25113 wild-type, Δ*pta*, and Δ*ackA* cells grown in glucose media were compared by western blotting (Figure 2B). The acetylation level of ThrRS was aligned with the amount of intracellular AcP, confirming AcP-dependent acetylation of ThrRS in cells. We also compared the acetylation levels at exponential and stationary phases (Figure 2B). Clearly, cells at the stationary phase had significantly increased acetylation levels, indicating that nonenzymatic acetylation mainly happens at the later time of cell growth when proteins act as carbon sinks for AcP and AcCoA. Then, we measured the tRNA charging activities of native ThrRSs form these strains (Figure 2C). The ThrRS from Δ*ackA* cells had a ~15% decreased activity of that from wild-type cells, consisting with site-specific kinetic analyses which showed that acetylation at several lysine

residues decreased the tRNA charging activity (Table 4). The ThrRS from Δ*pta* cells had a similar activity with that from wild-type cells. Since Δ*pta* blocks AcP production for nonenzymatic acetylation, the remaining acetylation should be mostly from enzymatic acetylation. The results indicated that nonenzymatic acetylation is the major mechanism for ThrRS acetylation, which was consistent with a very recent proteomic study of *E. coli* KATs which showed that the acetylation of ThrRS is mostly enzyme-independent [57].

We further tested the site-specificities of AcP-dependent acetylation. We treated all the four AARSs studied in this work with 2 mM AcP *in vitro*, and used LC-MS/MS to detect the acetylation sites (Table S3). Beside the selected lysine residues for site-specific studies, we also detected a number of additional lysine residues to be acetylated. Then we compared our lists of *in vitro* AcP acetylation with those in the acetylome database [33]. The *in vitro* AcP treatment could not generate all the acetylated lysine residues that have been identified by proteomic studies, indicating that those lysine residues only in the acetylome database may need specific acetyltransferases or additional cofactors or coenzymes for acetylation. Interestingly, there are also a number of lysine acetylation only found in AcP treatment, such as K62 and K430 of CysRS, K167 of AspRS, K359 of HisRS, and K249 of ThrRS. These positions may have their specific deacetylation mechanisms or some factors such as interacting proteins which can protect them from being acetylated in cells.

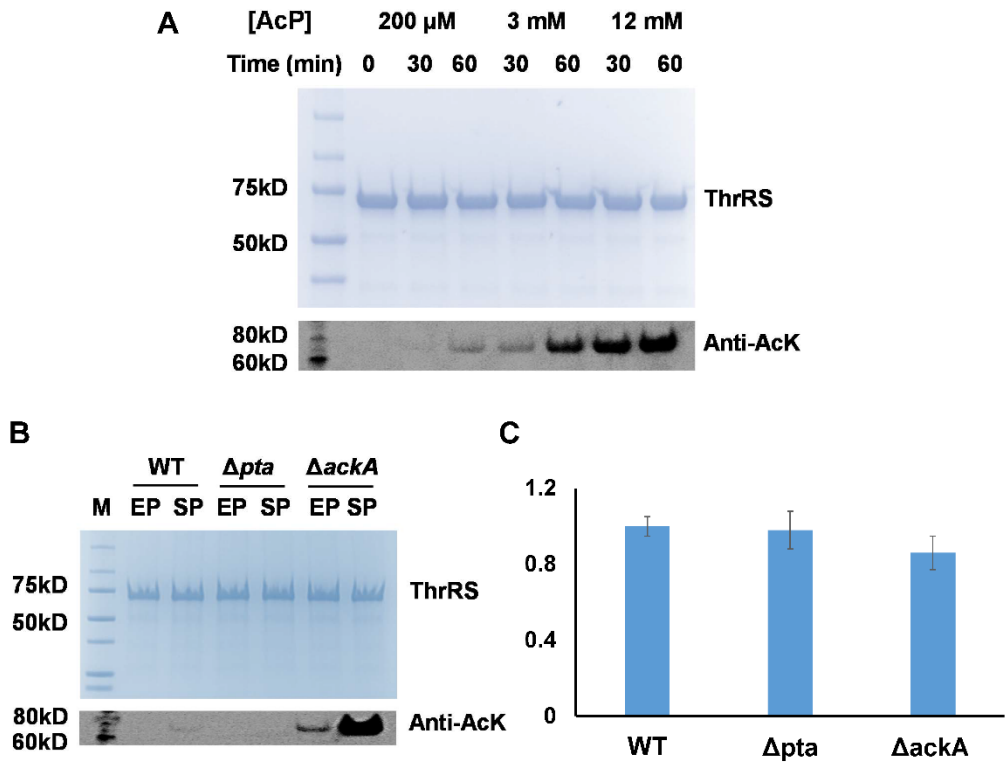


Figure 2.2 Acetylation of ThrRS by AcP. A) SDS-PAGE and western blotting analyses of purified ThrRS treated with different concentrations of AcP *in vitro*. B) SDS-PAGE and western blotting analyses of purified native ThrRSs from BW25113 wild-type, Δ*pta*, and Δ*ackA* cells at exponential (EP) and stationary phases (SP), respectively. The same amounts of proteins were loaded in all the experiments. C) The relative activities of tRNA charging of purified native ThrRSs from BW25113 wild-type, Δ*pta*, and Δ*ackA* cells. The tRNA charging activity of purified native ThrRSs from wild-type cells was set as 1.

2.4.4 The deacetylation process of *E. coli* ThrRS

Lysine deacetylases (KDACs) remove the acetyl group from the acetylated lysine residue in proteins. Classes I, II and IV KDACs are zinc-dependent amidohydrolases, while the class III KDAC (sirtuin) is NAD^+-dependent [58]. For each KDAC family, putative bacterial homologues have been identified, but CobB is still the only known KDAC in *E. coli* [59]. Thus, we tested the deacetylase activity of CobB for ThrRS both *in vivo* and *in vitro*.

For *in vivo* experiments, the native ThrRS was purified from the BW25113 wild-type or ΔcobB cells grown in M9 media with 0.2% glucose. Their acetylation levels were compared by western blotting (Figure 3A). The deletion of *cobB* increased the acetylation level of ThrRS significantly, indicating that CobB could deacetylate ThrRS *in vivo*. In this case, the similar acetylation levels of ThrRS in wild-type and Δ*yfiQ* strains could be caused by CobB deacetylation (Figure S37A). So, we determined the acetylation level of the native ThrRS from Δ*yfiQ* Δ*cobB* cells (Figure 3A). Again, deletion of *yfiQ* did not decrease the acetylation level of ThrRS with the Δ*cobB* background, confirming that it is not the KAT for ThrRS. Moreover, we compared the acetylation levels of the native ThrRS from Δ*ackA* Δ*cobB* cells and that from Δ*cobB* cells (Figure 3A). Clearly, Δ*ackA* Δ*cobB* cells had an increased acetylation, confirming that AcP-dependent nonenzymatic acetylation is the major acetylation mechanism for ThrRS.

To study the site-specificity of CobB for ThrRS, site-specifically acetylated ThrRS variants produced by the genetic incorporation system were treated with purified CobB for 1 hour, respectively. The acetylation levels of CobB-treated variants were determined by western blotting (Figure 3B). Most of the variants could be deacetylated completely by CobB-treatment

for 1 hour, but three CobB-resistant sites, K197, K200, and K226, were also identified. CobB prefers acetylated lysine residues which are at protein surface, since it needs to approach the residue first [60]. This could explain the CobB-resistance of K197 and K200 which are orientated toward inside of the protein (Figure S22A). However, K226 is located at the linker helix and faces outside in the crystal structure. One possible reason is that the orientation and location of K226 may not be fixed due to the flexibility of the linker helix [45].

Several proteomic studies on *E. coli* cells with the Δ*cobB* background showed increased acetylation stoichiometry in AARSs [36, 60]. However, studies have also shown that Δ*cobB* cells had higher AcP concentration to increase global acetylation [34]. Thus, although the deletion of the *cobB* gene increased the acetylation level of AARSs, we cannot conclude that CobB can deacetylate AARSs directly. By site-specific studies, previous studies and this work confirmed the CobB deacetylation activity on site-specifically acetylated AARS variants [15-17]. Moreover, we did find several CobB-resistant AARS variants such as K144 of TyrRS and K197, K200, and K226 of ThrRS, indicating there may be additional deacetylases or deacetylation mechanisms for those positions. Due to the limited knowledge of acetyltransferases and deacetylases in bacteria, de novo approaches to screen for additional modifying enzymes for AARS acetylation and deacetylation are necessary.

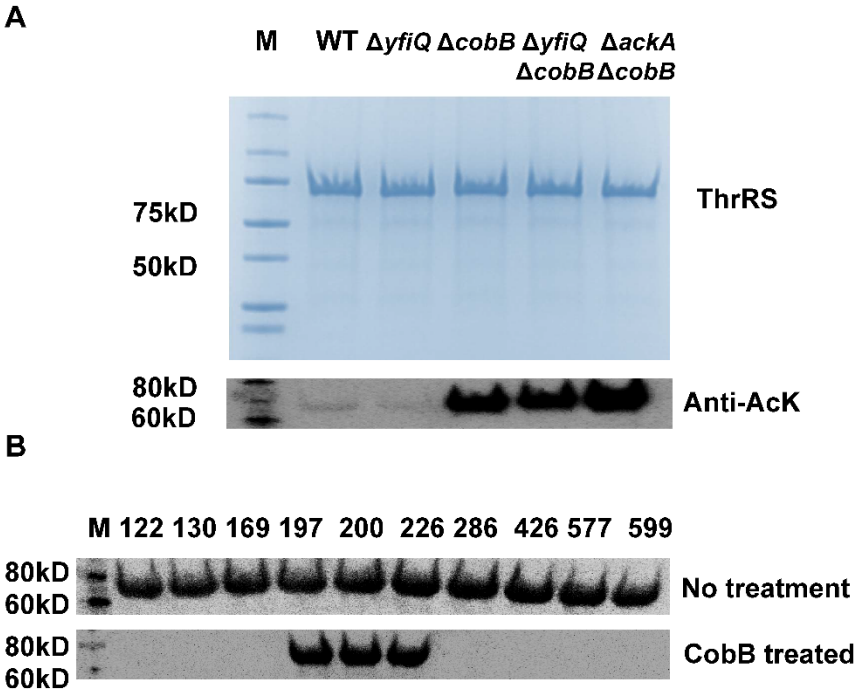


Figure 2.3 Deacetylation of ThrRS by CobB. A) SDS-PAGE and western blotting analyses of purified native ThrRS from BW25113 wild-type, Δ*yfiQ*, Δ*cobB, and* Δ*yfiQ* Δ*cobB, and* Δ*ackA* Δ*cobB* cells. B) Western blotting of site-specifically acetylated ThrRS variants treated with CobB for 1 hour. The same amounts of proteins were loaded in all the experiments. SDS-PAGE analyses of CobB-treated ThrRS variants was shown in Figure S39.

2.5 References

1. Ibba, M. and D. Soll, *Aminoacyl-tRNA synthesis.* Annu Rev Biochem, 2000. **69**: p. 617-50.

2. Ling, J., N. Reynolds, and M. Ibba, *Aminoacyl-tRNA synthesis and translational quality control.* Annu Rev Microbiol, 2009. **63**: p. 61-78.

3. Guo, M., X.L. Yang, and P. Schimmel, *New functions of aminoacyl-tRNA synthetases beyond translation.* Nat Rev Mol Cell Biol, 2010. **11**(9): p. 668-74.

4. Smirnova, E.V., et al., *Noncanonical functions of aminoacyl-tRNA synthetases.* Biochemistry (Mosc), 2012. **77**(1): p. 15-25.

5. Kim, D., N.H. Kwon, and S. Kim, *Association of aminoacyl-tRNA synthetases with cancer.* Top Curr Chem, 2014. **344**: p. 207-45.

6. Schwenzer, H., et al., *Pathogenic implications of human mitochondrial aminoacyl-tRNA synthetases.* Top Curr Chem, 2014. **344**: p. 247-92.

7. Sissler, M., L.E. Gonzalez-Serrano, and E. Westhof, *Recent Advances in Mitochondrial Aminoacyl-tRNA Synthetases and Disease.* Trends Mol Med, 2017. **23**(8): p. 693-708.

8. Lv, P.C. and H.L. Zhu, *Aminoacyl-tRNA synthetase inhibitors as potent antibacterials.* Curr Med Chem, 2012. **19**(21): p. 3550-63.

9. Yao, P. and P.L. Fox, *Aminoacyl-tRNA synthetases in medicine and disease.* EMBO Mol Med, 2013. **5**(3): p. 332-43.

10. Dewan, V., J. Reader, and K.M. Forsyth, *Role of aminoacyl-tRNA synthetases in infectious diseases and targets for therapeutic development.* Top Curr Chem, 2014. **344**: p. 293-329.

11. Pang, Y.L., K. Poruri, and S.A. Martinis, *tRNA synthetase: tRNA aminoacylation and beyond.* Wiley Interdiscip Rev RNA, 2014. **5**(4): p. 461-80.

12. Krishna, R.G. and F. Wold, *Post-translational modification of proteins.* Adv Enzymol Relat Areas Mol Biol, 1993. **67**: p. 265-98.

13. Arif, A., et al., *Two-site phosphorylation of EPRS coordinates multimodal regulation of noncanonical translational control activity.* Mol Cell, 2009. **35**(2): p. 164-80.

14. Kim, D.G., et al., *Interaction of two translational components, lysyl-tRNA synthetase and p40/37LRP, in plasma membrane promotes laminin-dependent cell migration.* FASEB J, 2012. **26**(10): p. 4142-59.

15. Venkat, S., et al., *Biochemical Characterization of the Lysine Acetylation of Tyrosyl-tRNA Synthetase in Escherichia coli.* Chembiochem, 2017. **18**(19): p. 1928-1934.

16. Ye, Q., et al., *Acetylation of lysine -amino groups regulates aminoacyl-tRNA synthetase activity in Escherichia coli.* J Biol Chem, 2017. **292**(25): p. 10709-10722.

17. Umehara, T., et al., *Lysine Acetylation Regulates Alanyl-tRNA Synthetase Activity in Escherichia coli.* Genes (Basel), 2018. **9**(10): p. e473.

18. Cao, X., et al., *Acetylation promotes TyrRS nuclear translocation to prevent oxidative damage.* Proc Natl Acad Sci U S A, 2017. **114**(4): p. 687-692.

19. Venkat, S., et al., *Characterizing Lysine Acetylation of Isocitrate Dehydrogenase in Escherichia coli.* J Mol Biol, 2018. **430**(13): p. 1901-1911.

20. Neumann, H., S.Y. Peak-Chew, and J.W. Chin, *Genetically encoding N^{ε}-acetyllysine in recombinant proteins.* Nat. Chem. Biol., 2008. **4**(4): p. 232-4.

21. Baba, T., et al., *Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection.* Mol Syst Biol, 2006. **2**: p. 2006.0008.

22. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6640-5.

23. Kitagawa, M., et al., *Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research.* DNA Res, 2005. **12**(5): p. 291-9.

24. Gan, Q., et al., *Expanding the genetic code of Salmonella with non-canonical amino acids.* Sci Rep, 2016. **6**: p. 39920.

25. Venkat, S., et al., *Genetically Incorporating Two Distinct Post-translational Modifications into One Protein Simultaneously.* ACS Synth Biol, 2018. **7**(2): p. 689-695.

26. Gan, Q. and C. Fan, *Increasing the fidelity of noncanonical amino acid incorporation in cell-free protein synthesis.* Biochim Biophys Acta Gen Subj, 2017. **1861**(11 Pt B): p. 3047-3052.

27. Fan, C., K. Ip, and D. Soll, *Expanding the genetic code of Escherichia coli with phosphotyrosine.* FEBS Lett, 2016. **590**(17): p. 3040-7.

28. Venkat, S., et al., *Genetically encoding thioacetyl-lysine as a non-deacetylatable analog of lysine acetylation in Escherichia coli.* FEBS Open Bio, 2017. **7**(11): p. 1805-1814.

29. Eriani, G., et al., *Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs.* Nature, 1990. **347**(6289): p. 203-6.

30. Ribas de Pouplana, L. and P. Schimmel, *Two classes of tRNA synthetases suggested by sterically compatible dockings on tRNA acceptor stem.* Cell, 2001. **104**(2): p. 191-3.

31. Burbaum, J.J. and P. Schimmel, *Structural relationships and the classification of aminoacyl-tRNA synthetases.* J Biol Chem, 1991. **266**(26): p. 16965-8.

32. Giege, R. and M. Springer, *Aminoacyl-tRNA Synthetases in the Bacterial World.* EcoSal Plus, 2016. **7**(1): p. doi:10.1128/ecosalplus.ESP-0002-2016.

33. Xu, H., et al., *PLMD: An updated data resource of protein lysine modifications.* J Genet Genomics, 2017. **44**(5): p. 243-250.

34. Baeza, J., et al., *Stoichiometry of site-specific lysine acetylation in an entire proteome.* J Biol Chem, 2014. **289**(31): p. 21326-38.

35. Meyer, J.G., et al., *Quantification of Lysine Acetylation and Succinylation Stoichiometry in Proteins Using Mass Spectrometric Data-Independent Acquisitions (SWATH).* J Am Soc Mass Spectrom, 2016. **27**(11): p. 1758-1771.

36. Weinert, B.T., et al., *Accurate Quantification of Site-specific Acetylation Stoichiometry Reveals the Impact of Sirtuin Deacetylase CobB on the E. coli Acetylome.* Mol Cell Proteomics, 2017. **16**(5): p. 759-769.

37. Hauenstein, S., et al., *Shape-selective RNA recognition by cysteinyl-tRNA synthetase.* Nat Struct Mol Biol, 2004. **11**(11): p. 1134-41.

38. Venkat, S., et al., *A facile protocol to generate site-specifically acetylated proteins in Escherichia coli.* J Vis Exp, 2017. **130**: p. e57061.

39. Zhang, C.M., et al., *Zinc-mediated amino acid discrimination in cysteinyl-tRNA synthetase.* J Mol Biol, 2003. **327**(5): p. 911-7.

40. Eiler, S., et al., *Synthesis of aspartyl-tRNA(Asp) in Escherichia coli--a snapshot of the second step.* EMBO J, 1999. **18**(22): p. 6532-41.

41. Thompson, D., et al., *Probing electrostatic interactions and ligand binding in aspartyl-tRNA synthetase through site-directed mutagenesis and computer simulations.* Proteins, 2008. **71**(3): p. 1450-60.

42. Arnez, J.G., et al., *The first step of aminoacylation at the atomic level in histidyl-tRNA synthetase.* Proc Natl Acad Sci U S A, 1997. **94**(14): p. 7144-9.

43. Tian, Q., et al., *Structural basis for recognition of G-1-containing tRNA by histidyl-tRNA synthetase.* Nucleic Acids Res, 2015. **43**(5): p. 2980-90.

44. Arnez, J.G., et al., *Crystal structure of histidyl-tRNA synthetase from Escherichia coli complexed with histidyl-adenylate.* EMBO J, 1995. **14**(17): p. 4143-55.

45. Sankaranarayanan, R., et al., *The structure of threonyl-tRNA synthetase-tRNA(Thr) complex enlightens its repressor activity and reveals an essential zinc ion in the active site.* Cell, 1999. **97**(3): p. 371-81.

46. Dock-Bregeon, A.C., et al., *Achieving error-free translation; the mechanism of proofreading of threonyl-tRNA synthetase at atomic resolution.* Mol Cell, 2004. **16**(3): p. 375-86.

47. Dock-Bregeon, A., et al., *Transfer RNA-mediated editing in threonyl-tRNA synthetase. The class II solution to the double discrimination problem.* Cell, 2000. **103**(6): p. 877-84.

48. Ling, J. and D. Soll, *Severe oxidative stress induces protein mistranslation through impairment of an aminoacyl-tRNA synthetase editing site.* Proc Natl Acad Sci U S A, 2010. **107**(9): p. 4028-33.

49. Roth, S.Y., J.M. Denu, and C.D. Allis, *Histone acetyltransferases.* Annu Rev Biochem, 2001. **70**: p. 81-120.

50. Starai, V.J. and J.C. Escalante-Semerena, *Identification of the protein acetyltransferase (Pat) enzyme that acetylates acetyl-CoA synthetase in Salmonella enterica.* J Mol Biol, 2004. **340**(5): p. 1005-12.

51. Paetz, W. and G. Nass, *Biochemical and immunological characterization of threonyl-tRNA synthetase of two borrelidin-resistant mutants of Escherichia coli K12.* Eur J Biochem, 1973. **35**(2): p. 331-7.

52. Wagner, G.R. and R.M. Payne, *Widespread and enzyme-independent Nepsilon-acetylation and Nepsilon-succinylation of proteins in the chemical conditions of the mitochondrial matrix.* J Biol Chem, 2013. **288**(40): p. 29036-45.

53. Weinert, B.T., et al., *Acetyl-phosphate is a critical determinant of lysine acetylation in E. coli.* Mol Cell, 2013. **51**(2): p. 265-72.

54. Kuhn, M.L., et al., *Structural, kinetic and proteomic characterization of acetyl phosphate-dependent bacterial protein acetylation.* PLoS One, 2014. **9**(4): p. e94816.

55. Klein, A.H., et al., *The intracellular concentration of acetyl phosphate in Escherichia coli is sufficient for direct phosphorylation of two-component response regulators.* J Bacteriol, 2007. **189**(15): p. 5574-81.

56. Wolfe, A.J., *The acetate switch.* Microbiol Mol Biol Rev, 2005. **69**(1): p. 12-50.

57. Christensen, D.G., et al., *Identification of Novel Protein Lysine Acetyltransferases in Escherichia coli.* MBio, 2018. **9**(5).

58. Hildmann, C., D. Riester, and A. Schwienhorst, *Histone deacetylases--an important class of cellular regulators with a variety of functions.* Appl Microbiol Biotechnol, 2007. **75**(3): p. 487-97.

59. Zhao, K., X. Chai, and R. Marmorstein, *Structure and substrate binding properties of cobB, a Sir2 homolog protein deacetylase from Escherichia coli.* J Mol Biol, 2004. **337**(3): p. 731-41.

60. AbouElfetouh, A., et al., *The E. coli sirtuin CobB shows no preference for enzymatic and nonenzymatic lysine acetylation substrate sites.* Microbiologyopen, 2015. **4**(1): p. 66-83.

CHAPTER III: A Synthetic Reporter for Probing Mistranslation in Living Cells

Hao Chen[1], Carson Ercanbrack[2], Tony Wang[3], Qinglei Gan[2], Chenguang Fan[1,2*]

[1]Cell and Molecular Biology Program, University of Arkansas, Fayetteville, AR, USA

[2]Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, AR, USA

[3]Depratment of Biology, University of Arkansas, Fayetteville, AR, USA

3.1 Abstract

Aminoacyl-tRNA synthetases (AARSs) play key roles in maintaining high fidelity of protein synthesis. They charge cognate tRNAs with corresponding amino acids, and hydrolyze mischarged tRNAs by editing mechanisms. Impairment of AARS editing activities can reduce the accuracy of tRNA aminoacylation to produce mischarged tRNAs, which cause mistranslation and cell damages. To evaluate the mistranslation rate of threonine codons in living cells, in this study we designed a quantitative reporter derived from the green fluorescent protein (GFP). The original GFP has multiple threonine codons which could affect the accuracy of measurement, so we generated a GFP variant containing only one threonine residue to specifically quantify mistranslation at the threonine codon. To validate, we applied this single-threonine GFP reporter to evaluate mistranslation at the threonine codon with mutations or modifications of threonine-tRNA synthetase and compared with other methods of mistranslation evaluation, which showed that this reporter is reliable and facile to use.

3.2 Introduction

From genetic information stored in DNA to functional proteins, there are many processes including DNA replication, transcription, and translation. All these steps can produce mistakes. But the error rates for each step are quite different, ranging from 10^{-8} during DNA replication [1] to 10^{-4} during protein synthesis [2]. Although there are many mechanisms during translation to maintain high fidelity of protein synthesis, the error rate of translation is still relatively high which makes protein mistranslation a remarkable research topic in the fields of biochemistry [3-5].

Protein mistranslation brings non-cognate amino acids locally into one specific codon or globally replaces one amino acid with another regardless of codons to generate a series of protein variants which can have impaired protein functions or produce aggregation due to protein misfolding [6]. Thus traditionally, protein mistranslation is thought to be harmful or lethal to cells. Indeed, it causes a wide range of human diseases including neurological disorders, developmental disorders, viral infections, and cancers [7-10]. However, very recently, a number of studies have shown that protein mistranslation is not always detrimental [11-13]. Misincorporation of methionine can protect cells from reactive oxygen species [14]. Mischaring of tRNA can mask amino acid starvation to alter stress response signaling [15, 16]. Protein mistranslation is also found to be involved in pathogenicity by increasing the antigenic diversity of pathogens to bypass host immune defenses or promoting phenotypic heterogeneity to increase opportunities to survive under different stress conditions [17-19].

To better study protein mistranslation, several methods have been developed to detect or quantify misincorporation of non-cognate amino acids [6, 11]. The most sensitive and accurate approach is based on mass spectrometry (MS) [20]. However, MS cannot be used for in vivo studies and depends on high-resolution MS facilities which could not be easily accessible for all the research groups in this field. To be conveniently used in living cells, several reporters has been designed derived from the green fluorescent protein (GFP) including eGFP T65V [21], GFP Y66K [22], GFP E222Q [23], and eGFP D129P variants [24]. But all these variants have multiple target amino acids at other sites (For example, in eGFP T65V, there are 18 valine residues at other positions), which could confound quantification results. To overcome this issue, in this study, we designed a GFP variant with only one threonine residue to quantify mistranslation of the threonine codon in living cells more precisely. This strategy can be applied to develop different single-target amino acid GFP variant reporters for quantifying mistranslation of specific codons.

3.3 Materials and Methods

3.3.1 Construction of sfGFP variants

The gene of Thr-free sfGFP was ordered from Integrated DNA Technologies (Coralville, IA, USA). The sequence is in Supplementary Material. The primers for generating sfGFP variants with single threonine residue are listed in Table S1. The mutations were made by the Q5 Site-Directed Mutagenesis Kit from New England Biolabs (Ipswich, MA, USA) following the manufacturer's protocol. The DNA sequences of these variants were confirmed by DNA

sequencing (Eurofins Genomics, Louisville, KY, USA). The gene of sfGFP or its variants was transformed into *E. coli* BL21 DE3 cells (New England Biolabs, Ipswich, MA, USA) for expression. Detailed procedures for vector construction were in Supplementary Materials.

3.3.2 Fluorescence reading

The strain harboring the plasmid to express WT sfGFP or its variants was inoculated into 2 mL minimal medium (M9 medium and 0.4% glucose) with 100 μg/mL streptomycin and incubated at 37 °C overnight, individually. 10 μL of each overnight culture was added into 190 μL fresh minimal medium with 100 μg/mL streptomycin to OD_{600nm} ~ 0.15. The 200 μL mixture was pipetted into a well in a 96-well plate with supplementary 0.1 mM IPTG to induce the protein expression. The fluorescence intensity (excitation wavelength 485/20 nm; emission wavelength 528/20 nm) and cell growth (OD_{600nm}) were monitored by a microplate-reader with continuous shaking at 37 °C for 6 hours. Means and standard deviations were calculated by three replicates for initial screening and five replicates for measuring mistranslation rates. Normalized fluorescence was the fluorescence reading divided by the cell culture density (OD_{600nm}) or by the protein yield quantified by ELISA. The mistranslation rate equals 1 minus the rate of non-mistranslation which is the normalized fluorescence of cells expressing threonine-tRNA synthetase (ThrRS) variants divided by that of cells expressing WT-ThrRS.

3.3.3 Protein expression, purification, and characterization

For easy purification, a His_6-tag fused to the WT sfGFP or its variants by PCR reaction used the NEBuilder HiFi DNA Assembly Master Mix Kit. The expression and purification procedures followed the previous protocol [25]. Detailed procedures were in Supplementary Materials. The site-specifically acetylated ThrRS variant was generated by our established genetic incorporation system [26]. Protein concentrations were measured by the Bradford Protein Assay (Bio-Rad, Hercules, CA, USA). Purified proteins were fractionated on a 12% SDS PAGE gel and visualized by the Bio-Safe Coomassie Stain (Bio-Rad). To quantify GFP yields, ELISA was performed with anti-His_6 tag (Abcam, Cambridge, MA, USA).

3.3.4 Mass and CD spectrometry

The LC-MS/MS analyses were performed by Yale University Keck Proteomics Facility following previous protocols [27]. The purified ThrRS was digested in gel by trypsin, and analyzed by LC-MS/MS on an LTQ Orbitrap XL equipped with a nanoACQUITY UPLC system. The Mascot search algorithm was used to search for the acetyllysine modifications. CD spectrometry was performed by previous protocols [28]. The CD spectra were recorded on a J-1500 CD Spectrometer. Purified proteins were diluted to a concentration of 0.1 mg/ml in 5 mM Tris-HCl pH 7.8, 0.1 M KCl, and scanned from 190 nm to 250 nm with a 20 nm/min speed. Scanning was performed three times for each sample and the average was plotted.

3.4 Results

3.4.1 Developing a precise reporter to evaluate mistranslation at the threonine codon *in vivo*

ThrRS uses a zinc ion to discriminate against the valine (without hydroxyl group) at the activation step and utilizes the N-terminal editing domain to hydrolyze mischarged $tRNA^{Thr}$ with serine (with a smaller size) [29, 30]. Because ThrRS solves such unique double discrimination problem with isosteric amino acids, it has attracted many attentions for studying its mistranslation [31]. However, there is still no facile method to quantify mistranslation of threonine codons in living cells. In this study, we engineered the super-folder GFP (sfGFP) for this purpose. As mentioned, ThrRS can mischarge $tRNA^{Thr}$ with serine. Besides serine, alanyl-tRNA synthetase (AlaRS) can also mischarge $tRNA^{Thr}$ with alanine [32]. The sfGFP has 18 threonine residues in total. All these threonine codons could be mistranslated as serine or alanine. Thus, we observe the overall effect of mistranslation of all these threonine codons on GFP fluorescence. Mistranslation at some sites could increase fluorescence, while that at others could impair fluorescence. So, we may underestimate or overestimate the actual mistranslation. To eliminate possible interference from mistranslation at other threonine codons and focus on one specific threonine codon, we aimed to generate a sfGFP variant containing only one threonine residue which is essential for its fluorescence (Figure 3.1).

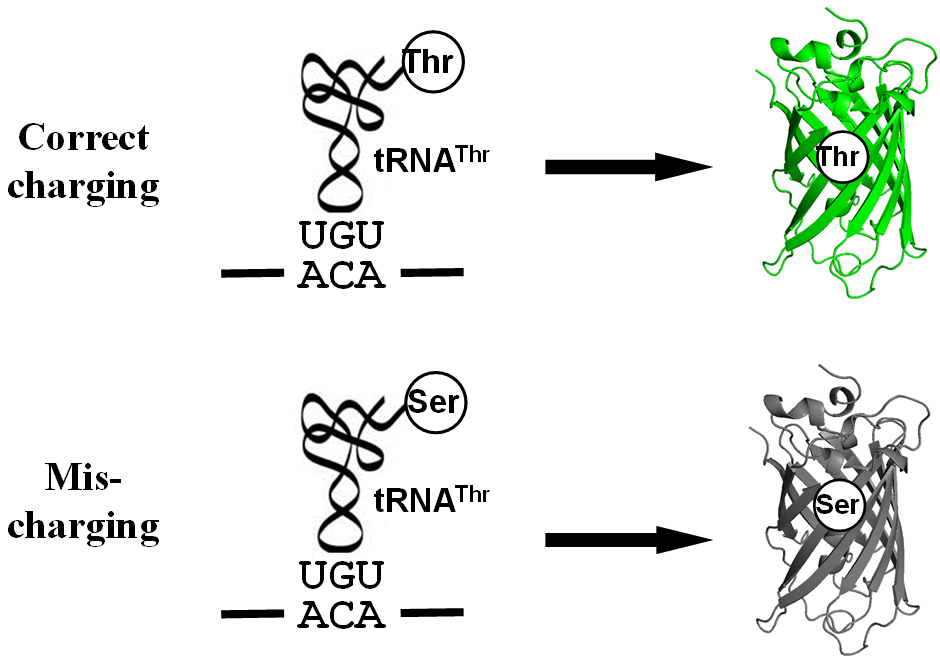


Figure 3.1 A scheme for the Thr-dependent sfGFP variant. The misincorporation of serine at the threonine codon due to mischarged Ser-tRNA^{Thr} eliminates the fluorescence of sfGFP.

First, we mutated all the threonine codons in the gene of sfGFP to serine codons because serine is the most common mistranslated amino acid for threonine codons. As expected, this Thr-free sfGFP variant (TF-sfGFP) had no fluorescence (Figure 2). The protein yield of TF-sfGFP was similar to that of WT-sfGFP with the same expression condition (Figure S1), indicating that substitution of threonine residues with serine at all 18 sites does not affect GFP expression significantly. We also performed CD spectrometry analyses with both TF-sfGFP and WT-sfGFP to see whether such substitution can affect GFP folding. The result showed that there was no significant difference between them (Figure S2). In the next step, we put the threonine codon back to its original position in the gene of TF-sfGFP individually to generate 18 single-threonine sfGFP variants in total. We measured fluorescence intensities for these sfGFP variants (Figure 2). Among these variants, the TF-sfGFP T203 variant restored ~ 20% fluorescence with one threonine residue alone. We also noted that protein yields of all these variants were similar to that of WT-sfGFP with the same expression condition (Figure S1).

T203 plays a critical role in forming the structure similar to a proton pump for GFP and is well conserved among GFP-like proteins [33], which explained the recovery of sfGFP fluorescence. Besides T203, T65 is well known to increase the fluorescence of GFP [34]. But the TF-sfGFP T65 variant in this study had no significant fluorescence, probably because it is a good enhancer, but not essential for producing fluorescence. Actually, the original amino acid at this site in WT-GFP is serine. Besides serine, AlaRS can also mischarge $tRNA^{Thr}$ with alanine [32]. To eliminate the possibility that the substitution of threonine with alanine at position 203 can also generate fluorescence, we generated the TF-sfGFP A203 variant. This variant had no

fluorescence (Figure 3.2). Thus, the TF-sfGFP T203 variant can be used as the reporter for quantifying mistranslation of the threonine codon, because i) mistranslation of the threonine codon to serine or alanine will eliminate the fluorescence of sfGFP; ii) there is no interference of mistranslation of other threonine codons in the sfGFP; iii) the fluorescence intensity depends on the ratio of mistranslated sfGFP and total amounts of sfGFP.

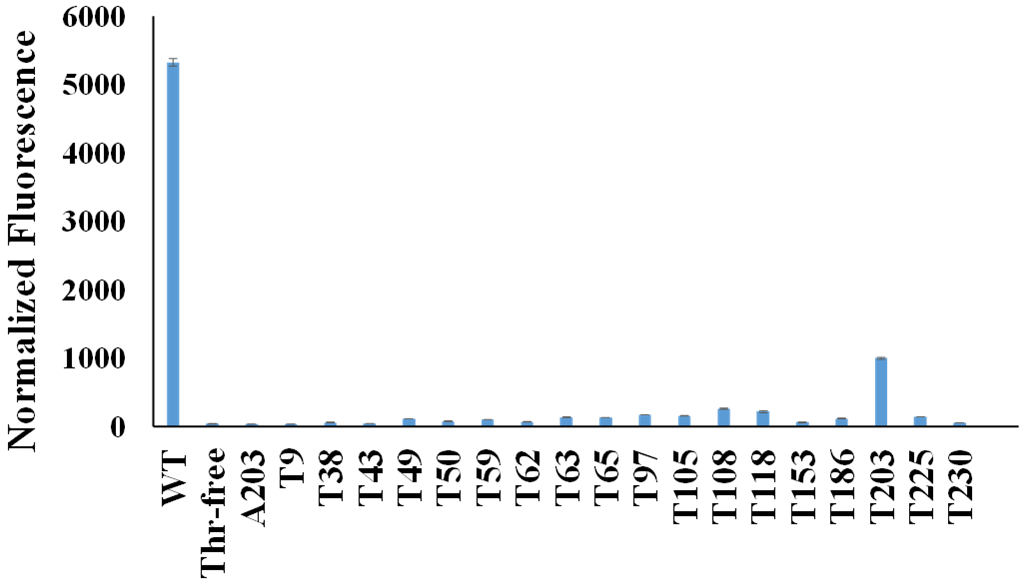


Figure 3.2 Normalized fluorescence intensities of cells expressing sfGFP variants. Both the fluorescence intensity and the cell culture density of each strain expressing different sfGFP variants were monitored. The fluorescence intensity at 6 hour after induction for each strain was normalized by its cell culture density (OD_{600nm}). The fluorescence intensity of the strain harboring the empty vector (pCDF-1b) was used as the baseline which was subtracted from each reading. Mean and standard deviations were calculated based on three replicates.

3.4.2 Testing the reporter for mistranslation of threonine codons

In the previous quantitative MS study on ThrRS-mediated mistranslation, the ThrRS editing-deficient variant ThrRS C182A was used to evaluate the mistranslation rate at threonine codons [20]. To compare our method with the MS approach, we expressed the TF-sfGFP T203 variant in the strain containing ThrRS C182A. The strain containing WT-ThrRS was used as the control (Table 3.1). Both the protein yield and cell culture density were decreased in the strain containing ThrRS C182A, which is consistent with the previous study [35]. To address potential effects of *E. coli* cell auto-fluorescence on GFP fluorescence readings which could be caused by different cell densities and stress conditions [36, 37], we used the corresponding stains harboring the same vector but without the reporter gene as controls. Fluorescence readings were subtracted with corresponding backgrounds. Moreover, we compared two approaches to normalize fluorescence intensities, either by cell culture densities from OD_{600nm} (Figure S3) or by protein yields from ELISA. Both normalization approaches had similar results and showed that the mistranslation rate caused by ThrRS C182A variant is ~3%, which is consistent with previous quantitative MS studies [20]. As cell culture densities were monitored simultaneously with fluorescence reading, so we used this more facile approach for later experiments.

Table 3.1 Comparison of two normalization methods for fluorescence intensities.

	WT-ThrRS	ThrRS C182A
Fluorescence reading	1256 ± 17	979 ± 12
sfGFP yield by ELISA (g/L)	0.402 ± 0.013	0.323 ± 0.016
Normalized Fluorescence by ELISA*	3124 ± 14	3031 ± 10
Mistranslation rate by ELISA**	0	3.1 ± 0.6 %
Cell culture density (OD_{600nm})	0.668 ± 0.019	0.538 ± 0.013
Normalized Fluorescence by OD_{600nm}	1880 ± 11	1820 ± 12
Mistranslation rate by OD_{600nm}	0	3.2 ± 0.9 %

***** Normalized fluorescence was fluorescence readings subtracted with corresponding backgrounds of no insert vector control and divided by the sfGFP yield quantified by ELISA or the cell culture density by OD_{600nm} at 6 hours after induction. Mean and standard deviations were calculated based on five replicates. Mean and standard deviations for normalized fluorescence were calculated based on values of normalized fluorescence for each sample rather than normalizing averaged raw fluorescence reading by averaged ELISA or OD values.

** The mistranslation rate equals 1 minus the rate of non-mistranslation which is the normalized fluorescence of cells expressing the ThrRS C182A variant divided by that of cells expressing WT-ThrRS.

To compare TF-sfGFP T203 and WT-sfGFP in evaluating mistranslation, we tested the effect of the ThrRS C182A variant on their fluorescence, individually. We also applied the similar strategy in previous mistranslation studies [38] and added serine in growth media to force threonine to serine mistranslation (Figure 3.3). With the increase of serine concentrations, both reporters had increased mistranslation rates as expected. WT-sfGFP gave higher mistranslation rates than the single-threonine sfGFP reporter at all conditions, probably because threonine to serine substitution at other threonine sites further decreased its fluorescence. In this case, we could overestimate mistranslation rates by using WT-sfGFP as the reporter. Furthermore, single-threonine sfGFP reporter was less sensitive to serine concentrations than WT-sfGFP, which is another advantage of this reporter.

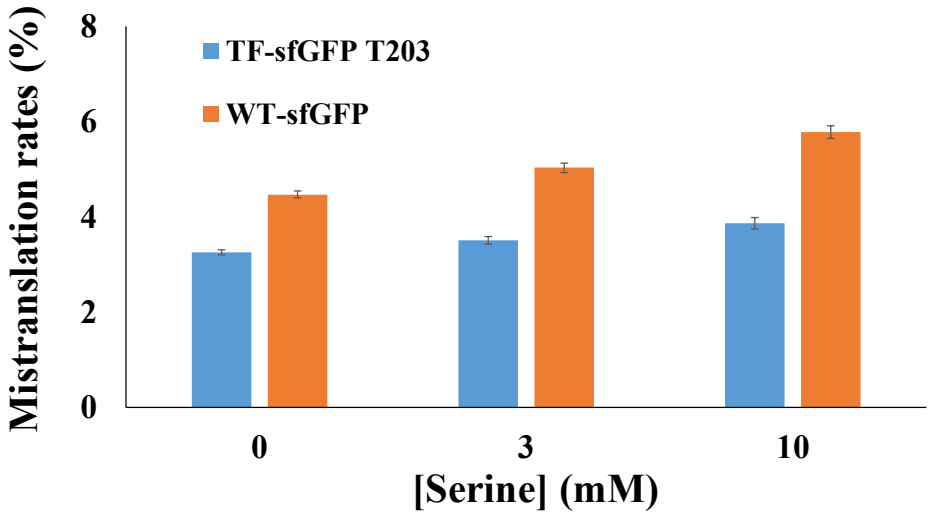


Figure 3.3 The effect of serine concentrations on sfGFP reporters. Mistranslation rates were calculated with normalized fluorescence by cell culture densities. Mean and standard deviations were calculated based on five replicates.

In our recent study on acetylation of ThrRS, we found that acetylation of K169 in *E. coli* ThrRS can generate Ser-mischarged $tRNA^{Thr}$ *in vitro* [28]. However, the effect of acetylation of ThrRS on mistranslation of threonine codons in living cells is unknown. So, we applied the single-threonine sfGFP reporter to evaluate the impact of acetylation of ThrRS on threonine mistranslation in vivo. The site-specifically acetylated *E. coli* ThrRS at K169 (ThrRS-169AcK) was generated by the genetic code expansion strategy in E. coli cells and confirmed by LC-MS/MS (Figure S4). The TF-sfGFP T203 variant was co-expressed in the strain containing ThrRS 169AcK. Results showed that the mistranslation rate caused by ThrRS acetylation was 4.13 ± 0.06 %. K169 is located at the opening of the editing site of ThrRS (Figure S5). We proposed that the ThrRS-169AcK variant has an impaired activity to hydrolyze mischarged Ser-$tRNA^{Thr}$ due to steric hindrance from the additional acetyl group. The crystallography study on this acetylated ThrRS variant is ongoing.

3.5 Discussion

Although several GFP variants have been developed to estimate protein mistranslation in living cells [21-24], they cannot precisely quantify mistranslation rates. Take the eGFP T65V as an example, it was made to evaluate valine to threonine mistranslation by an editing defective ValRS in mammalian cells [38]. There are 18 other valine residues in the eGFP. When those valine residues are substituted with threonine, the folding or fluorescence properties could also be altered. In this case, the mistranslation rate could be underestimated because valine to threonine substitution at other positions can decrease its fluorescence. Actually, authors in that study realized

this problem, and used a 10-fold of free threonine concentration in the medium to force the valine to threonine mistranslation at all valine codons. They noticed a decreased mistranslation rate (from 16.7% to 14.4%), and concluded that this number 'loosely' reflects the degree of mistranslation in cells [38].

In this study, we generated a sfGFP variant with only one threonine codon, so there is no interference from other positions, which makes this strategy unique. This strategy could also be applied in other AARS-mediated mistranslation studies. For the example of ValRS mischarging studies, a Val-free GFP variant could be generated by replacing valine resides with isosteric amino acids such as alanine. By searching for the GFP-like protein data bank [33], there is no conserved valine residue in GFP, so it is possible to generate a Val-free GFP variant without eliminating fluorescence. Then the Val-free GFP with T65V substitution could be used for precisely quantify the mistranslation rate at valine codons.

3.6 References

1. Kunkel, T.A. and K. Bebenek, *DNA replication fidelity.* Annu Rev Biochem, 2000. **69**: p. 497-529.

2. Ellis, N. and J. Gallant, *An estimate of the global error frequency in translation.* Mol Gen Genet, 1982. **188**(2): p. 169-72.

3. Evans, C.R., et al., *Errors during Gene Expression: Single-Cell Heterogeneity, Stress Resistance, and Microbe-Host Interactions.* MBio, 2018. **9**(4).

4. Hoffman, K.S., P. O'Donoghue, and C.J. Brandl, *Mistranslation: from adaptations to applications.* Biochim Biophys Acta Gen Subj, 2017. **1861**(11 Pt B): p. 3070-3080.

5. Schwartz, M.H. and T. Pan, *Function and origin of mistranslation in distinct cellular contexts.* Crit Rev Biochem Mol Biol, 2017. **52**(2): p. 205-219.

6. Mohler, K. and M. Ibba, *Translational fidelity and mistranslation in the cellular response to stress.* Nat Microbiol, 2017. **2**: p. 17117.

7. Kapur, M. and S.L. Ackerman, *mRNA Translation Gone Awry: Translation Fidelity and Neurological Disease.* Trends Genet, 2018. **34**(3): p. 218-231.

8. Ou, X., et al., *Errors in translational decoding: tRNA wobbling or misincorporation?* PLoS Genet, 2019. **15**(3): p. e1008017.

9. Lant, J.T., et al., *Pathways to disease from natural variations in human cytoplasmic tRNAs.* J Biol Chem, 2019. **294**(14): p. 5294-5308.

10. Schimmel, P., *Development of tRNA synthetases and connection to genetic code and disease.* Protein Sci, 2008. **17**(10): p. 1643-52.

11. Ribas de Pouplana, L., et al., *Protein mistranslation: friend or foe?* Trends Biochem Sci, 2014. **39**(8): p. 355-62.

12. Steiner, R.E. and M. Ibba, *Regulation of tRNA-dependent translational quality control.* IUBMB Life, 2019. **71**(8): p. 1150-1157.

13. Pan, T., *Adaptive translation as a mechanism of stress response and adaptation.* Annu Rev Genet, 2013. **47**: p. 121-37.

14. Lee, J.Y., et al., *Promiscuous methionyl-tRNA synthetase mediates adaptive mistranslation to protect cells against oxidative stress.* J Cell Sci, 2014. **127**(Pt 19): p. 4234-45.

15. Bullwinkle, T.J. and M. Ibba, *Translation quality control is critical for bacterial responses to amino acid stress.* Proc Natl Acad Sci U S A, 2016. **113**(8): p. 2252-7.

16. Mohler, K., et al., *Editing of misaminoacylated tRNA controls the sensitivity of amino acid stress responses in Saccharomyces cerevisiae.* Nucleic Acids Res, 2017. **45**(7): p. 3985-3996.

17. Ackermann, M., *A functional perspective on phenotypic heterogeneity in microorganisms.* Nat Rev Microbiol, 2015. **13**(8): p. 497-508.

18. Miranda, I., et al., *Candida albicans CUG mistranslation is a mechanism to create cell surface variation.* MBio, 2013. **4**(4).

19. Li, L., et al., *Naturally occurring aminoacyl-tRNA synthetases editing-domain mutations that cause mistranslation in Mycoplasma parasites.* Proc Natl Acad Sci U S A, 2011. **108**(23): p. 9378-83.

20. Mohler, K., et al., *MS-READ: Quantitative measurement of amino acid incorporation.* Biochim Biophys Acta Gen Subj, 2017. **1861**(11 Pt B): p. 3081-3088.

21. Nangle, L.A., C.M. Motta, and P. Schimmel, *Global effects of mistranslation from an editing defect in mammalian cells.* Chem Biol, 2006. **13**(10): p. 1091-100.

22. Biddle, W., M.A. Schmitt, and J.D. Fisk, *Evaluating Sense Codon Reassignment with a Simple Fluorescence Screen.* Biochemistry, 2015. **54**(50): p. 7355-64.

23. Su, H.W., et al., *The essential mycobacterial amidotransferase GatCAB is a modulator of specific translational fidelity.* Nat Microbiol, 2016. **1**(11): p. 16147.

24. Hoffman, K.S., et al., *Genetic selection for mistranslation rescues a defective co-chaperone in yeast.* Nucleic Acids Res, 2017. **45**(6): p. 3407-3421.

25. Venkat, S., et al., *Genetically encoding thioacetyl-lysine as a non-deacetylatable analog of lysine acetylation in Escherichia coli.* FEBS Open Bio, 2017. **7**(11): p. 1805-1814.

26. Venkat, S., et al., *A Facile Protocol to Generate Site-Specifically Acetylated Proteins in Escherichia Coli.* J Vis Exp, 2017(130).

27. Venkat, S., et al., *Characterizing lysine acetylation of Escherichia coli type II citrate synthase.* FEBS J, 2019. **286**(14): p. 2799-2808.

28. Chen, H., et al., *Site-Specifically Studying Lysine Acetylation of Aminoacyl-tRNA Synthetases.* ACS Chem Biol, 2019. **14**(2): p. 288-295.

29. Sankaranarayanan, R., et al., *Zinc ion mediated amino acid discrimination by threonyl-tRNA synthetase.* Nat Struct Biol, 2000. **7**(6): p. 461-5.

30. Dock-Bregeon, A.C., et al., *Achieving error-free translation; the mechanism of proofreading of threonyl-tRNA synthetase at atomic resolution.* Mol Cell, 2004. **16**(3): p. 375-86.

31. Dock-Bregeon, A., et al., *Transfer RNA-mediated editing in threonyl-tRNA synthetase. The class II solution to the double discrimination problem.* Cell, 2000. **103**(6): p. 877-84.

32. Sun, L., et al., *Evolutionary Gain of Alanine Mischarging to Noncognate tRNAs with a G4:U69 Base Pair.* J Am Chem Soc, 2016. **138**(39): p. 12948-12955.

33. Ong, W.J., et al., *Function and structure of GFP-like proteins in the protein data bank.* Mol Biosyst, 2011. **7**(4): p. 984-92.

34. Heim, R., A.B. Cubitt, and R.Y. Tsien, *Improved green fluorescence.* Nature, 1995. **373**(6516): p. 663-4.

35. Ling, J. and D. Söll, *Severe oxidative stress induces protein mistranslation through impairment of an aminoacyl-tRNA synthetase editing site.* Proc Natl Acad Sci U S A, 2010. **107**(9): p. 4028-33.

36. Surre, J., et al., *Strong increase in the autofluorescence of cells signals struggle for survival.* Sci Rep, 2018. **8**(1): p. 12088.

37. Mihalcescu, I., et al., *Green autofluorescence, a double edged monitoring tool for bacterial growth and activity in micro-plates.* Phys Biol, 2015. **12**(6): p. 066016.

38. Lee, J.W., et al., *Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration.* Nature, 2006. **443**(7107): p. 50-5.

CHAPTER IV: Conclusion and future directions

The canonical function of aminoacyl-tRNA synthetases (AARSs) is to charge the tRNAs with cognate amino acids, which is a critical step during ribosomal protein synthesis. The fidelity of the faithful translation process is based on the accurate recognition of amino acids and tRNA in AARSs. Although, synthetases have already been equipped with so many strong recognition mechanisms, extra proof-reading or editing functions are still needed for some AARSs which usually have the tendency to mischarge the tRNA in their catalytic domains. Regardless of certain beneficial examples, mistranslation caused by mischarged tRNAs usually can cause damage to cells, thus leading to disease conditions. Recently, lysine acetylation has been implied to have significant effects on the catalytic activities of AARSs, however, it is just a start, more explorations are needed to understand the impacts of lysine acetylation on these synthetases.

The major focus of this study is to broaden the understanding of how acetylation can affect the functions of AARSs. Here, the site-specific effects of lysine acetylation on four AARSs in *E. coli*, including class I cysteinyl-tRNA synthetase (CysRS) as well as class II aspartyl-tRNA synthetase (AspRS), histidyl-tRNA synthetase (HisRS) and threonyl-tRNA synthetase (ThrRS), are studied.

To achieve the site-specific incorporation of lysine acetylation into these AARSs, an optimized co-translational acetyllysine (AcK) incorporation system (AcKRS) was established and applied. To be specific, an engineered orthogonal pyrrolysyl-tRNA synthetase could correctly recognize AcK and charged it to the cognate $tRNA^{pyl}$, then the acetyllysyl- $tRNA^{pyl}$ would be translated into specific position of target protein in response the stop codon UAG. With

this system and according to our criterion for acetylation sites selection (have the highest acetylation stoichiometry *in vivo*; identified in more than three proteomic studies), this study generated 25 site-specifically acetylated variants of the four selected synthetases.

For class I CysRS, K73, K76, K175, K269, and K310 were selected for site-specific acetylation studies. Among them, K175 was at CP domain which helped stabilize the tRNA. The acetylation occurring in this site could decrease the tRNA charging rate. K269 was at class I-conserved KMSKS motif (part of catalytic domain). Its acetylation could cause decreasing of both cysteine activation and tRNA charging. The acetylation of K310 at anticodon binding domain could cause a slight decrease of tRNA affinity.

For class II AspRS and HisRS, acetylation in most selected lysine sites showed no significant impact, except two sites, K81of AspRS and K375 of HisRS. K81 was at the anticodon recognition region of AspRS, whose acetylation have a slight impact on the tRNA binding. K375 in HisRS interacted with anticodon stem of tRNA, and its acetylation could cause a decreasing of tRNA charging rate. Interestingly, those selected residues close to active site, such as K283 of AspRS as well as K53 and K65 of HisRS, showed no detectable effects.

For class II ThrRS, ten lysine residues had been selected in this study. K122, K130, K169, K197 and K200 were at the editing domain, among which K200's acetylation could cause the decreasing of tRNA charging rate. The acetylation of K286 in class II-conserved motif 1 could decrease the tRNA aminoacylation rate. Like other AARSs in this study, acetylation occurring at tRNA anticodon loop binding region of ThrRS (K577 and K599) could decrease the tRNA affinity to enzyme. Additionally, the serine misactivation test was used to characterize the effect

of lysine acetylation on ThrRS editing function. Surprisingly, it was detected that the acetylation of K169 could impair the editing function of ThrRS and mischarge serine to $tRNA^{Thr}$. Since K169 is located at the opening of the editing domain, its acetylation was thought to affect the proper orientation of Ser-$tRNA^{Thr}$ for deacylation. Furthermore, both acetylation and deacetylation tests were conducted to understand the acetylation process of ThrRS, and the results demonstrated that the nonenzymatic acetylation and CobB-dependent deacetylation were preferred by this enzyme.

To develop a GFP-based biosensor for detection of mistranslation caused by ThrRS with deficient editing function, 18 single-Thr sfGFP variants were generated and the fluorescence intensities for each were measured. Among them, only the T-203 variant restored about 20% fluorescence, while others and Thr-free variant showed extremely weak or no fluorescence signal. Therefore, the T-203 variant was selected as the potential probe to quantify the ThrRS-related mistranslation rate. Here, this probe was used to detect the mistranslation caused by ThrRS with K169 acetylation, and the mistranslation rate was around 12%, which indicates that K169 acetylation could influence the editing function of ThrRS *in vivo*.

In summary, by exploring the site-specific effects of lysine acetylation on four AARSs whose acetylation hasn't been studied before, we will be able to expand the knowledge on how acetylation affects the functions of AARSs. With an emphasis on ThrRS and its editing function, this study firstly indicates that acetylation can impair the editing function of AARSs in *E. coli*. Although with this study, the effects of acetylation on total eight AARSs (four from class I and four from class II) has been investigated, the remaining AARSs still are needed to be studied.

And current studies focus more on the acetylation occurring in prokaryotic AARSs, however, the effects of acetylation on eukaryotic or even human AARSs may be different and need more attention in future. It is also remarkable that AARSs acetylation is indicated to be related to several cancers, but more future research is required to understand the relationship between them. Additionally, most of studies have been aware of the effects of lysine acetylation on catalytic activities of AARSs, whereas a few try to explore the potential influence of acetylation on noncanonical functions of AARSs. Furthermore, this study has developed a strategy to design the probe for mistranslation caused by ThrRS with impaired editing domain, which will be extended to quantify tRNA-mischarging by other AARSs in living cells for different purposes.

SUPPLEMENTARY DATA

Supplementary data for chapter II

Supplementary methods

General molecular biology

Protein concentrations were measured by the Bradford Protein Assay. The purified AARSs and their variants were fractionated on a 4-20% SDS-PAGE gel and visualized by the Bio-Safe Coomassie Stain. For western blotting, the fractionated SDS-PAGE gel was transferred onto a PVDF membrane by using the Trans-Blot Turbo Transfer System. The PVDF membrane was incubated at room temperature with gentle shacking in the blocking buffer (5% bovine serum albumin, 0.1% Tween 20 in Tris-buffered saline) for 2 h. The horseradish peroxidase (HRP)-conjugated acetyllysine antibody (Cell Signaling Technology) was diluted 1:1000 with the blocking buffer and soaked the blocked PVDF membrane overnight at 4 °C. The membrane was prepared for chemiluminescence detection by using Pierce ECL Western Blotting substrates (Thermo Scientific).

Expression and purification of acetylated AARS variants

The genes of AARSs and their acetylated variants were cloned into the pCDF-1b plasmid with a C-terminal His_6-tag, and transformed into BL21 (DE3) cells together with the pTech plasmid harboring the AcK incorporation system for expression. The expression strain was grown on 500 mL of LB medium supplemented with 100 μg/mL streptomycin and 50 μg/mL chloramphenicol at 37°C. For wild-type AARSs, protein expression was induced by adding 0.1

mM isopropyl β-D-1-thiogalactopyranoside (IPTG) to the culture when it reached an absorbance of 0.6 at 600 nm. For acetylated variant expression, 10 mM AcK and 30 mM nicotinamine (NAM) were added to the culture with an absorbance of 0.4 at 600 nm. After incubating for additional 1 hour, 0.1 mM IPTG was added to induce protein expression. Cells were incubated at 18°C for an additional 12 h and harvested by centrifugation at 5,000 × g for 10 min at 4 °C. Cells were washed with 20 mL of 25 mM Tris (pH 7.8) and 10 mM NaCl twice, centrifuged, and stored at 80°C. Frozen cells were suspended in 15 mL of lysis buffer which contains 50 mM Tris (pH 7.8), 300 mM NaCl, 20 mM imidazole, 20 mM NAM, and 5 mM β-mercaptoethanol with cocktail protease inhibitors, and broken by 40 kHz sonication at 70% power output with 5 cycles of 30 s short bursts, followed by intervals of 30 s for cooling to form crude extracts. The crude extract was centrifuged at 20,000 × g for 25 min at 4°C. The soluble fraction was filtered through a 0.45 μm membrane and loaded onto a column containing 2 mL of Ni-NTA resin previously equilibrated with 20 mL lysis buffer. The column was then washed with 20 mL of 50 mM Tris (pH 7.8), 300 mM NaCl, and 60 mM imidazole. The protein bound to the column was finally eluted with 2 mL of 50 mM Tris (pH 7.8), 300 mM NaCl, and 150 mM imidazole. The purified protein was dialyzed twice with 2 L of 25 mM Tris (pH 7.8), 10 mM NaCl, 1 mM dithiothreitol (DTT), and 50% glycerol for 3 hours, and stored at 80°C.

Mass spectrometry (MS) analyses

The full-length MS analyses were performed by the University of Arkansas State-wide Mass Spectrometry Facility. Purified proteins were dialyzed with pure water, diluted to 0.05

mg/mL, and analyzed by a Bruker UltraflexII TOF-TOF with a MALDI ionization source. Spectra were obtained in the positive ion, linear mode. The LC-MS/MS analyses were performed by Yale University Keck Proteomics Facility. Proteins were digested in gel by trypsin, and analyzed by LC-MS/MS on an LTQ Orbitrap XL equipped with a nanoACQUITY UPLC system. The Mascot search algorithm was used to search for the appropriate noncanonical substitution.

In vitro transcription and purification of tRNAs

The template plasmid containing the tRNA gene was purified with the plasmid maxi kit, digested with BstNI, purified, and resolved in water. The transcription reaction contained Tris (50 mM, pH 8), each of UTP, CTP, GTP, and ATP (5 mM, pH 7.0), MgCl2 (25 mM), spermidine (2 mM), DTT (10 mM), pyrophosphatase (0.5 μg/ml), BstNI digested DNA template (60 μg/ml), and T7 RNA polymerase (0.5 mg/ml). The reaction was performed in a reaction volume (40 ml) for overnight at 37°C. The tRNA was purified on 12% denaturing polyacrylamide gel containing 8M urea and TBE buffer [Tris (90 mM), boric acid (90 mM), and EDTA (2 mM)]. UV shadowing illuminated the pure tRNA band, which was excised and extracted three times with sodium acetate (1 M, pH 5.3) at 4°C. The tRNA extractions were then ethanol precipitated, dissolved in RNase-free distilled water, pooled, and finally desalted using a Biospin 30 column. The tRNA was refolded by heating to 100°C for 5 min and slow cooling to room temperature. At 65°C, $MgCl_2$ was added to a final concentration of 10 mM to aid folding.

In vitro acetylation and deacetylation assay

For acetylation, the reaction (100 μL) was performed in the mixture containing HEPES (50 mM, pH 7.0), EDTA (0.1 mM), DTT (2 mM), glycerol (10%), and sodium butyrate (10 mM). The acetylation was carried out by adding AARS (10 μg), YfiQ (10 μg), and acetyl-CoA (0.2 mM) or acetyl-phosphate only (varied concentrations). Reaction mixtures were completely mixed and incubated at 37 °C for 1 h. For deacetylation, the reaction (100 μL) was performed in the mixture containing HEPES-KOH (50 mM, pH 7.0), $MgCl_2$ (5 mM), NAD+ (1 mM), DTT (2 mM), glycerol (10%), acetylated AARS variant (10 μg), and CobB (10 μg), incubated at 37 °C for 1 h.

Table S1 The list of strains with plasmids used in this study.

Strain	Plasmid
BW25113	
BW25113 Δ*cobB*	
BW25113 Δ*yfiQ*	
BW25113 Δ*pta*	
BW25113 Δ*ackA*	
BW25113 Δ*yfiQ* Δ*cobB*	
BW25113 Δ*ackA* Δ*cobB*	
AG1	pCA24N-CobB
AG1	pCA24N-YfiQ
AG1	pCA24N-RcsB
BL21(DE3)	pCDF-CysRS
BL21(DE3)	pCDF-CysRS (73TAG) + pTech-AcK-OTS*
BL21(DE3)	pCDF-CysRS (76TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-CysRS (175TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-CysRS (269TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-CysRS (310TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-AspRS
BL21(DE3)	pCDF-AspRS (81TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-AspRS (186TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-AspRS (283TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-AspRS (332TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-AspRS (353TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-HisRS
BL21(DE3)	pCDF-HisRS (30TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-HisRS (53TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-HisRS (65TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-HisRS (189TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-HisRS (378TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-ThrRS
BL21(DE3)	pCDF-ThrRS (122TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-ThrRS (130TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-ThrRS (169TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-ThrRS (197TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-ThrRS (200TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-ThrRS (226TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-ThrRS (286TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-ThrRS (426TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-ThrRS (577TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-ThrRS (599TAG) + pTech-AcK-OTS
BL21(DE3)	pCDF-ThrRS (K156A)

*: OTS is the abbreviation for orthogonal translational modification.

Table S2 Kinetic parameters of misactivation of serine by ThrRS.

	Serine		
	*K*M (mM)	*k*cat (S^{-1})	*k*cat / *K*M (S^{-1} mM^{-1})
Wild-type ThrRS	62 ± 9	17 ± 4	0.27
122AcK	72 ± 11	16 ± 3	0.22
130AcK	73 ± 8	18 ± 2	0.25
169AcK	68 ± 4	16 ± 4	0.24
197AcK	69 ± 18	17 ± 1	0.21
200AcK	65 ± 7	16 ± 3	0.25
226AcK	63 ± 5	16 ± 2	0.25
286AcK	78 ± 21	15 ± 2	0.19
426AcK	75 ± 13	14 ± 3	0.19
577AcK	66 ± 8	16 ± 2	0.24
599AcK	63 ± 16	17 ± 3	0.27

* Kinetic parameters were determined by ATP-PPi-exchange assays. Mean and standard deviations were calculated from three biological replicates. The catalytic efficiency of threonine activation by wild-type ThrRS is 260 S^{-1} mM^{-1}.

Table S3 The list of acetylation sites of AARSs treated with AcP *in vitro* and those reported by proteomic studies of *E. coli* cells. Additional acetylated lysine residues by AcP treatmentcompared to selected sites in this study were marked in red.

	In vitro AcP treatment	Acetylome database
CysRS	K62, K73, K76, K175, K269, K310, K430	K3, K60, K73, K76, K98, K175, K185, K188, K269, K282, K310, K328
AspRS	K81, K167, K186, K283, K332, K353, K370, K400, K415	K58, K81, K143, K184, K186, K214, K260, K283, K315, K332, K342, K347, K353, K359, K370, K399, K415, K513
HisRS	K30, K53, K65, K189, K191, K211, K359, K370, K378	K3, K30, K53, K65, K189, K191, K197, K370, K371, K378, K400
ThrRS	K33, K122, K130, K169, K197, K200, K226, K227, K241, K249, K286, K314, K330, K426, K577, K599, K614, K627	K33, K122, K130, K169, K197, K200, K226, K227, K241, K284, K286, K294, K306, K314, K330, K346, K426, K570, K577, K599, K605, K614, K627

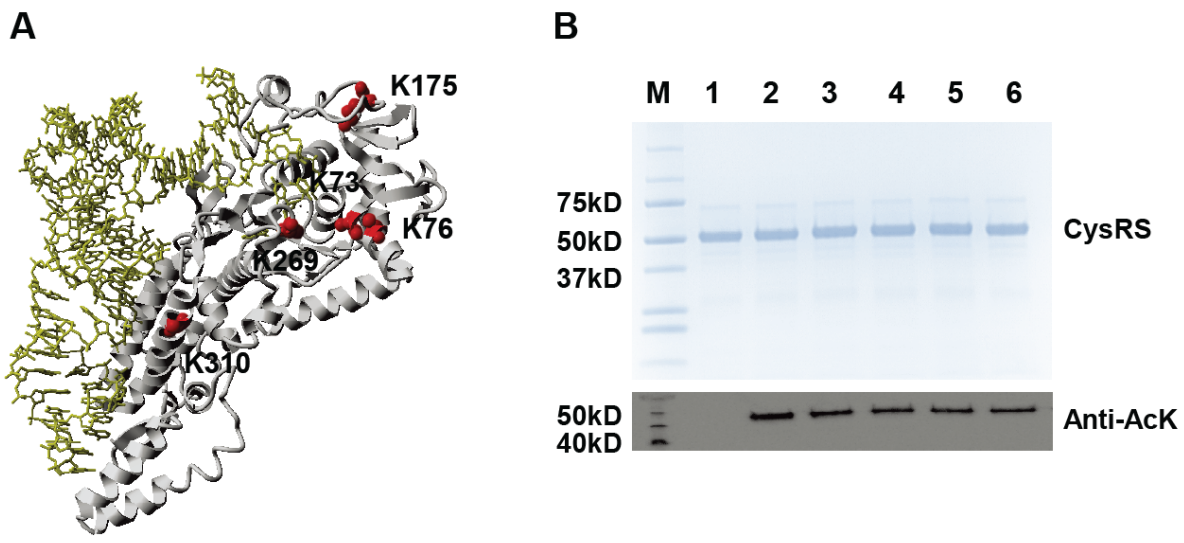


Figure S1 *E. coli* CysRS acetylation. A) Mapping of acetylated lysine residues on the crystal structure of CysRS (PDB ID: 1U0B). CysRS is in grey. $tRNA^{Cys}$ is in yellow. Selected residues arein red. B) SDS-PAGE and western blotting analyses of purified CysRS and its variants. Lane 1, wild-type CysRS; lane 2, CysRS-73AcK; lane 3, CysRS-76AcK; lane 4, CysRS-175AcK; lane 5; CysRS-269AcK; lane 6, CysRS-310AcK. The same amounts of proteins were loaded. Anti-AcK: the antibody of acetyllysine.

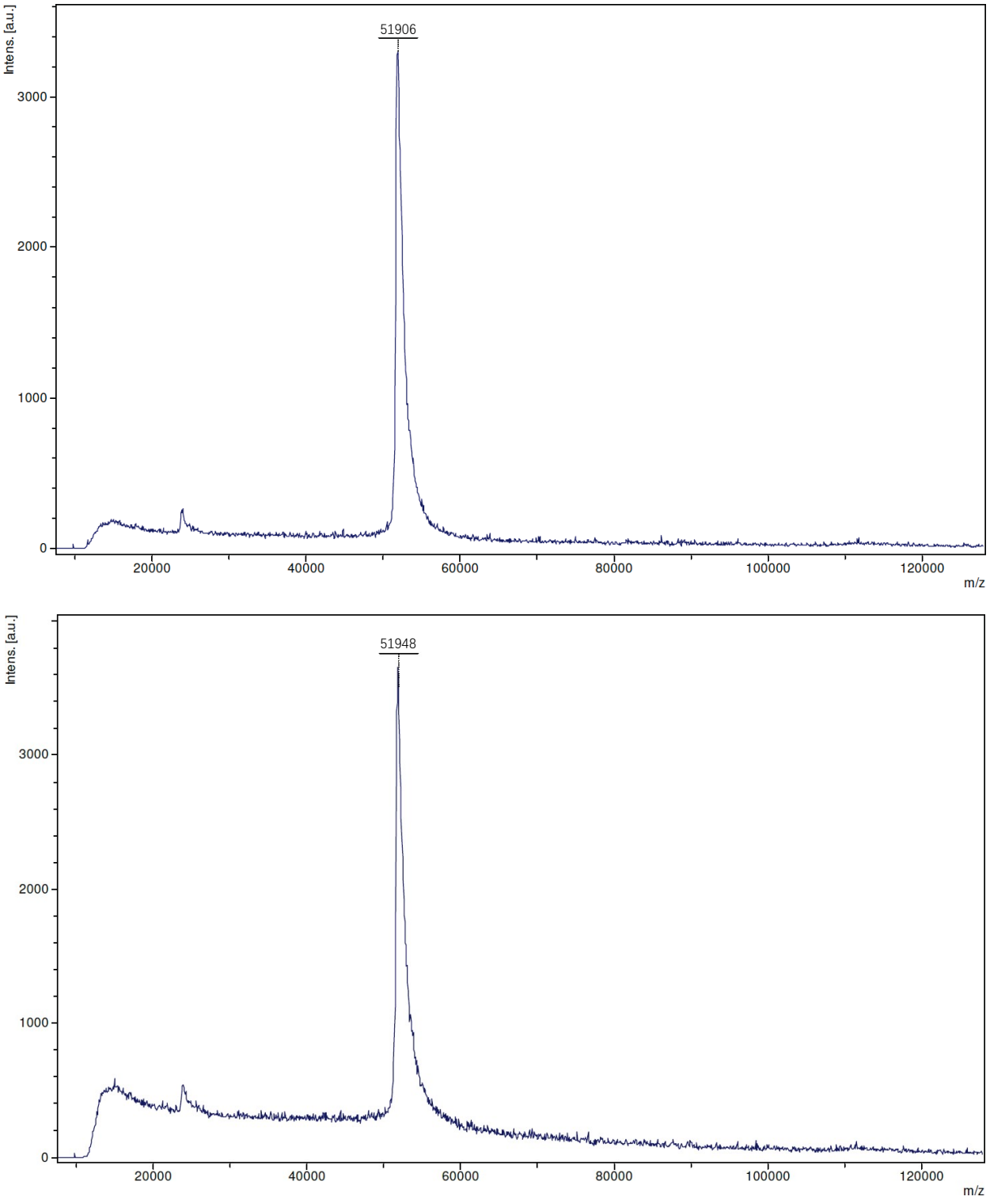
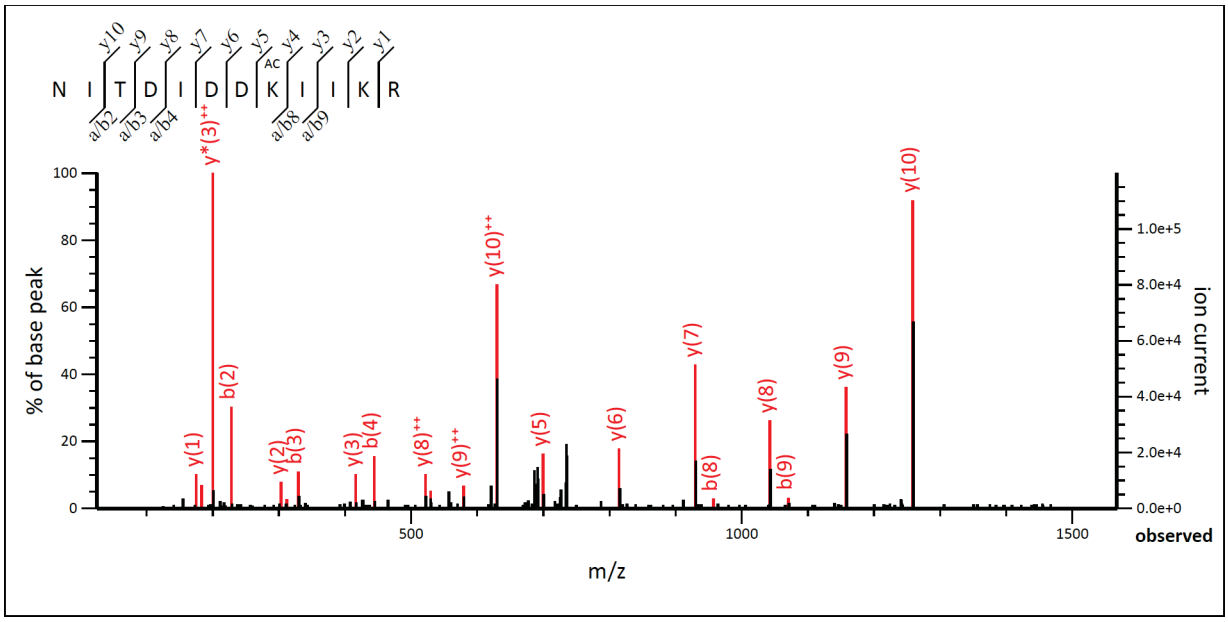
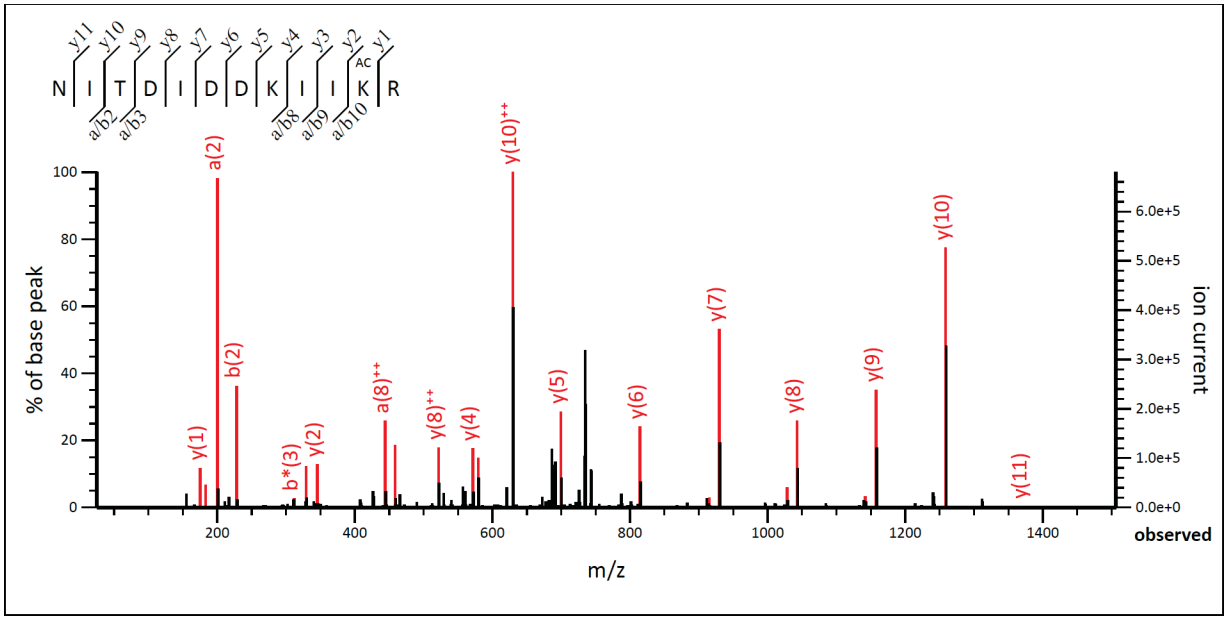


Figure S2 MALDI-TOF mass spectra of purified wild-type CysRS (upper, 51906 Da) and CysRS73-AcK (lower, 51948 Da) as a representative acetylated CysRS variant. The single peak of acetylated CysRS (ΔMW = 42 Da) indicated homogeneous incorporation of acetyllysine.



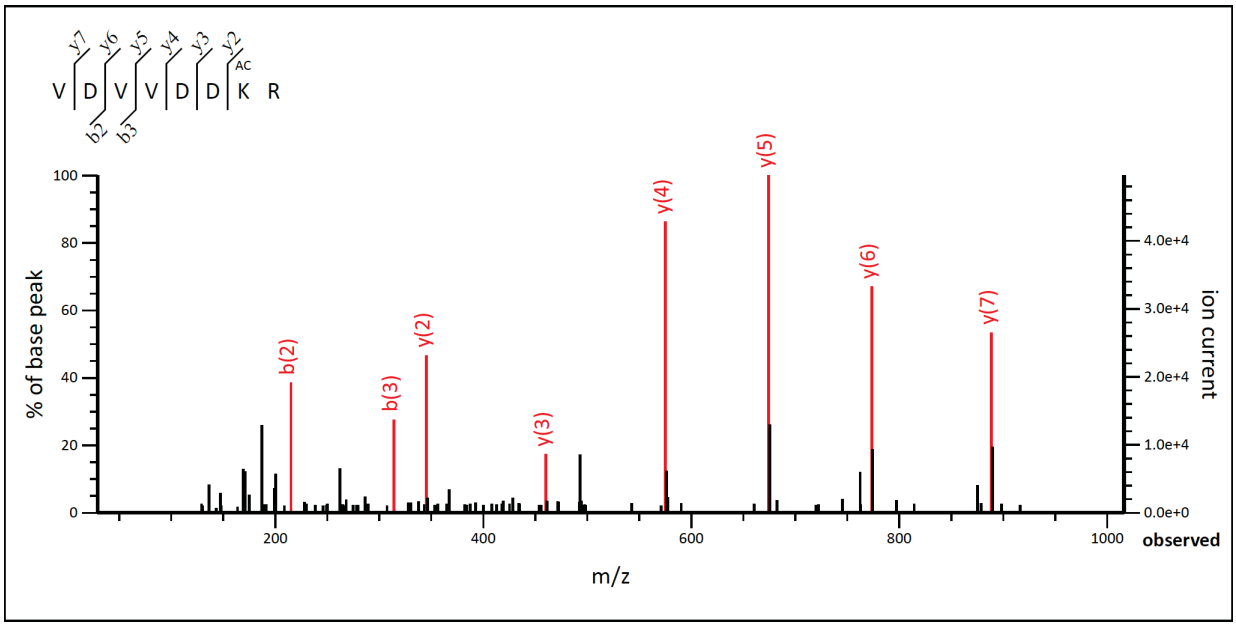
#	b	b++	Seq.	y	y++	#
1	115.0502	58.0287	N			12
2	**228.1343**	114.5708	I	1371.7842	686.3957	11
3	**329.1819**	165.0946	T	***1258.7001***	**629.8537**	10
4	**444.2089**	222.6081	D	***1157.6525***	**579.3299**	9
5	557.2930	279.1501	I	***1042.6255***	**521.8164**	8
6	672.3199	336.6636	D	***929.5415***	465.2744	7
7	787.3468	394.1771	D	***814.5145***	407.7609	6
8	**957.4524**	479.2298	AcK	***699.4876***	350.2474	5
9	**1070.5364**	535.7719	I	***529.3820***	265.1947	4
10	1183.6205	592.3139	I	***416.2980***	208.6526	3
11	1311.7155	656.3614	K	***303.2139***	152.1106	2
12			R	***175.1190***	88.0631	1

Figure S3 LC-MS/MS analysis of CysRS 73-AcK. The tandem mass spectrum of the peptide (residues 66-77) NITDIDDK^{AC}IIKR from purified CysRS 73-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated a, b or yion series. Matched peaks are in red.



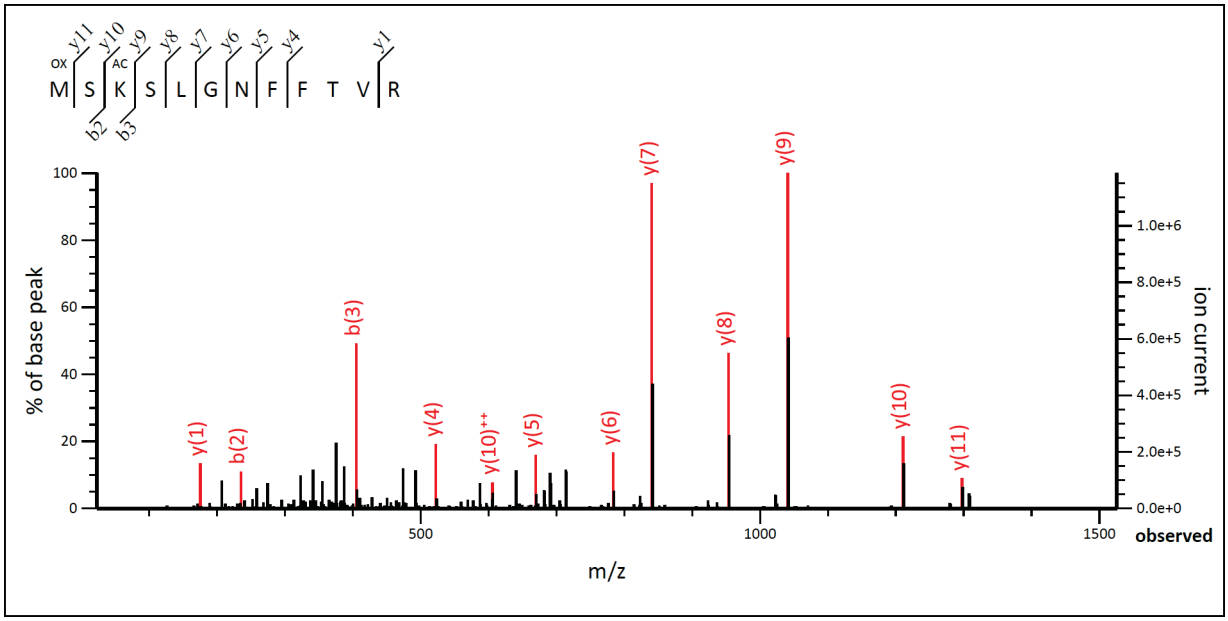
#	b	b++	Seq.	y	y++	#
1	115.0502	58.0287	N			12
2	**228.1343**	114.5708	I	*1371.7842*	686.3957	11
3	**329.1819**	165.0946	T	*1258.7001*	**629.8537**	10
4	**444.2089**	222.6081	D	*1157.6525*	**579.3299**	9
5	557.2930	279.1501	I	*1042.6255*	**521.8164**	8
6	672.3199	336.6636	D	*929.5415*	465.2744	7
7	787.3468	394.1771	D	*814.5145*	407.7609	6
8	**915.4418**	458.2245	K	*699.4876*	350.2474	5
9	**1028.5259**	514.7666	I	*571.3926*	286.1999	4
10	**1141.6099**	571.3086	I	*458.3085*	229.6579	3
11	1311.7155	656.3614	AcK	*345.2245*	173.1159	2
12			R	*175.1190*	88.0631	1

Figure S4 LC-MS/MS analysis of CysRS 76-AcK. The tandem mass spectrum of the peptide (residues 66-77) NITDIDDKIIK^{AC}R from purified CysRS 76-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated a, b or yion series. Matched peaks are in red.



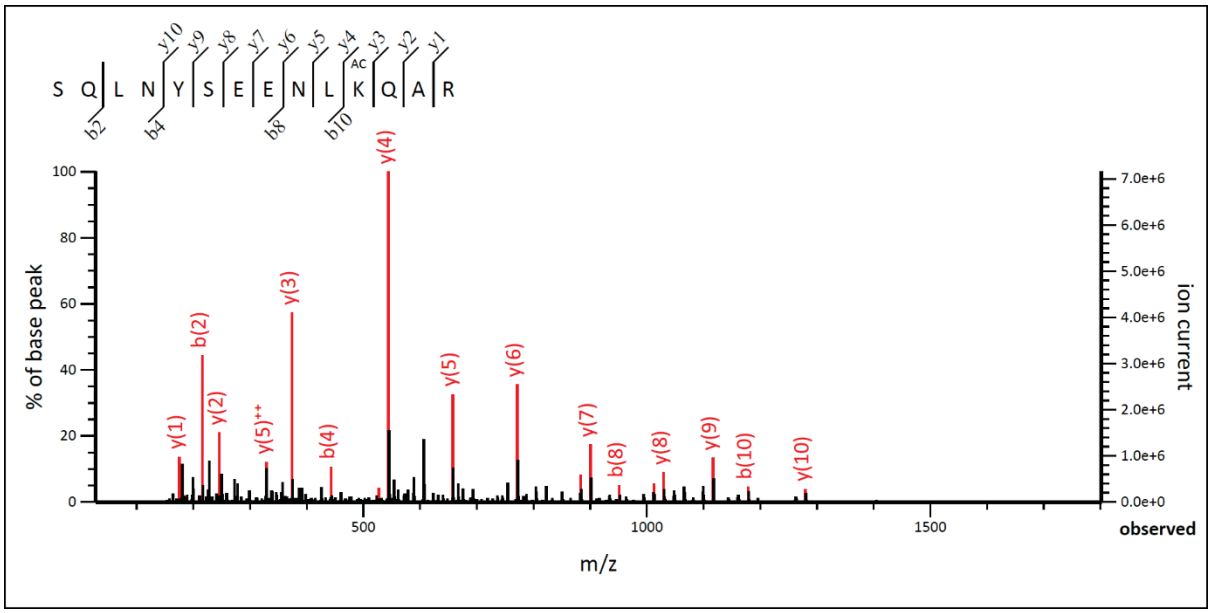
#	b	b^{++}	Seq.	y	y^{++}	#
1	100.0757	50.5415	V			8
2	**215.1026**	108.0550	D	***888.4421***	444.7247	7
3	**314.1710**	157.5892	V	***773.4152***	387.2112	6
4	413.2395	207.1234	V	***674.3468***	337.6770	5
5	528.2664	264.6368	D	***575.2784***	288.1428	4
6	643.2933	322.1503	D	***460.2514***	230.6293	3
7	813.3989	407.2031	AcK	***345.2245***	173.1159	2
8			R	175.1190	88.0631	1

Figure S5 LC-MS/MS analysis of CysRS 175-AcK. The tandem mass spectrum of the peptide (residues 169-176) VDVVDDK^{AC}R from purified CysRS 175-AcK. K^{AC} denotes AcK incorporation.The partial sequence of the peptide containing the AcK can be read from the annotated b or y ionseries. Matched peaks are in red.



#	b	b++	Seq.	y	y++	#
1	148.0427	74.5250	M			12
2	**235.0747**	118.0410	S	***1297.6899***	649.3486	11
3	**405.1802**	203.0938	K	***1210.6579***	**605.8326**	10
4	492.2123	246.6098	S	***1040.5524***	520.7798	9
5	605.2963	303.1518	L	***953.5203***	477.2638	8
6	662.3178	331.6625	G	***840.4363***	420.7218	7
7	776.3607	388.6840	N	***783.4148***	392.2110	6
8	923.4291	462.2182	F	***669.3719***	335.1896	5
9	1070.4975	535.7524	F	***522.3035***	261.6554	4
10	1171.5452	586.2762	T	375.2350	188.1212	3
11	1270.6136	635.8105	V	274.1874	137.5973	2
12			R	***175.1190***	88.0631	1

Figure S6 LC-MS/MS analysis of CysRS 269-AcK. The tandem mass spectrum of the peptide (residues 267-278) MSK^{AC}SLGNFFTVR from purified CysRS 269-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red.



#	b	b^{++}	Seq.	y	y^{++}	#
1	88.0393	44.5233	S			14
2	216.0979	108.5526	Q	1634.8133	817.9103	13
3	329.1819	165.0946	L	1506.7547	753.8810	12
4	443.2249	222.1161	N	1393.6706	697.3390	11
5	606.2882	303.6477	Y	1279.6277	640.3175	10
6	693.3202	347.1638	S	1116.5644	558.7858	9
7	822.3628	411.6850	E	1029.5323	515.2698	8
8	951.4054	476.2063	E	900.4898	450.7485	7
9	1065.4483	533.2278	N	771.4472	386.2272	6
10	1178.5324	589.7698	L	657.4042	329.2058	5
11	1348.6379	674.8226	AcK	544.3202	272.6637	4
12	1476.6965	738.8519	Q	374.2146	187.6110	3
13	1547.7336	774.3705	A	246.1561	123.5817	2
14			R	175.1190	88.0631	1

Figure S7 LC-MS/MS analysis of CysRS 310-AcK. The tandem mass spectrum of the peptide (residues 300-314) SQLNYSEENLK^{AC}QAR from purified CysRS 310-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red.

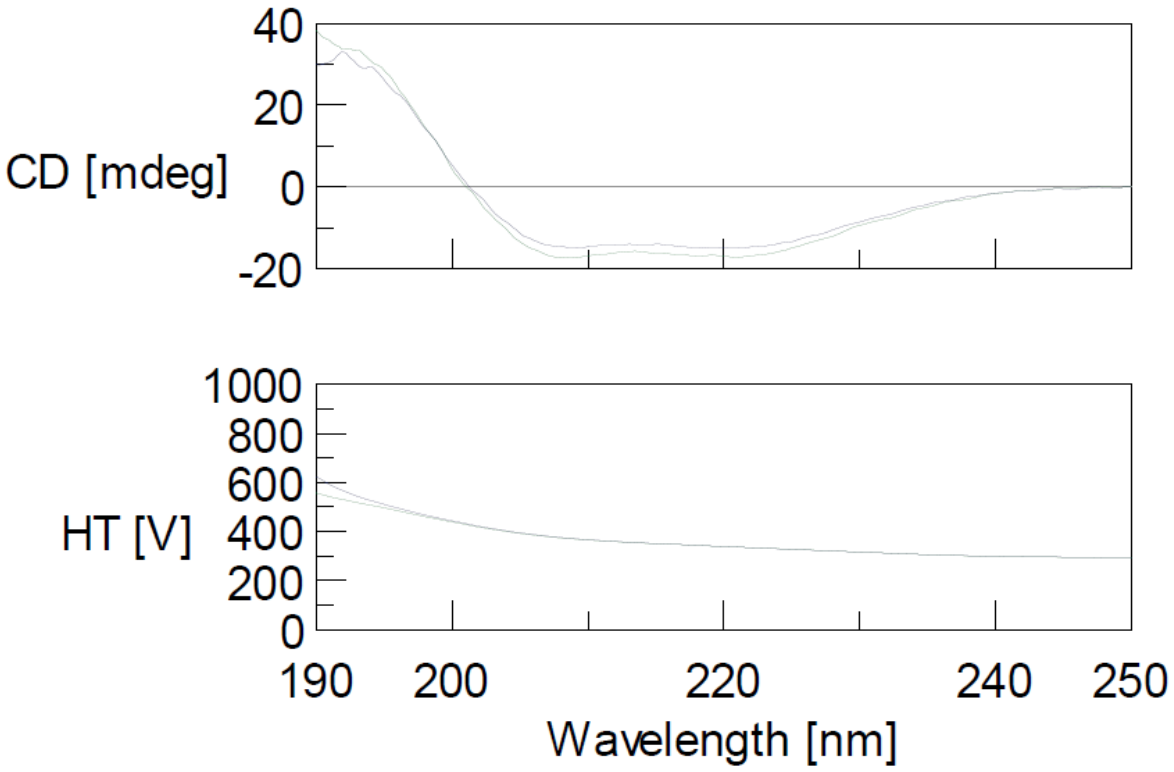


Figure S8 The circular dichroism (CD) spectra of the wild-type CysRS (blue lines) and its 269AcK variant (green lines). The CD spectra were recorded on a J-1500 CD Spectrometer. Purified enzymes were diluted to a concentration of 0.1 mg/ml in 5 mM Tris-HCl pH 7.8, 0.1 M KCl, and scanned from 190 nm to 250 nm with a 20 nm/min speed. Scanning was performed three times for each sample and the average was plotted.

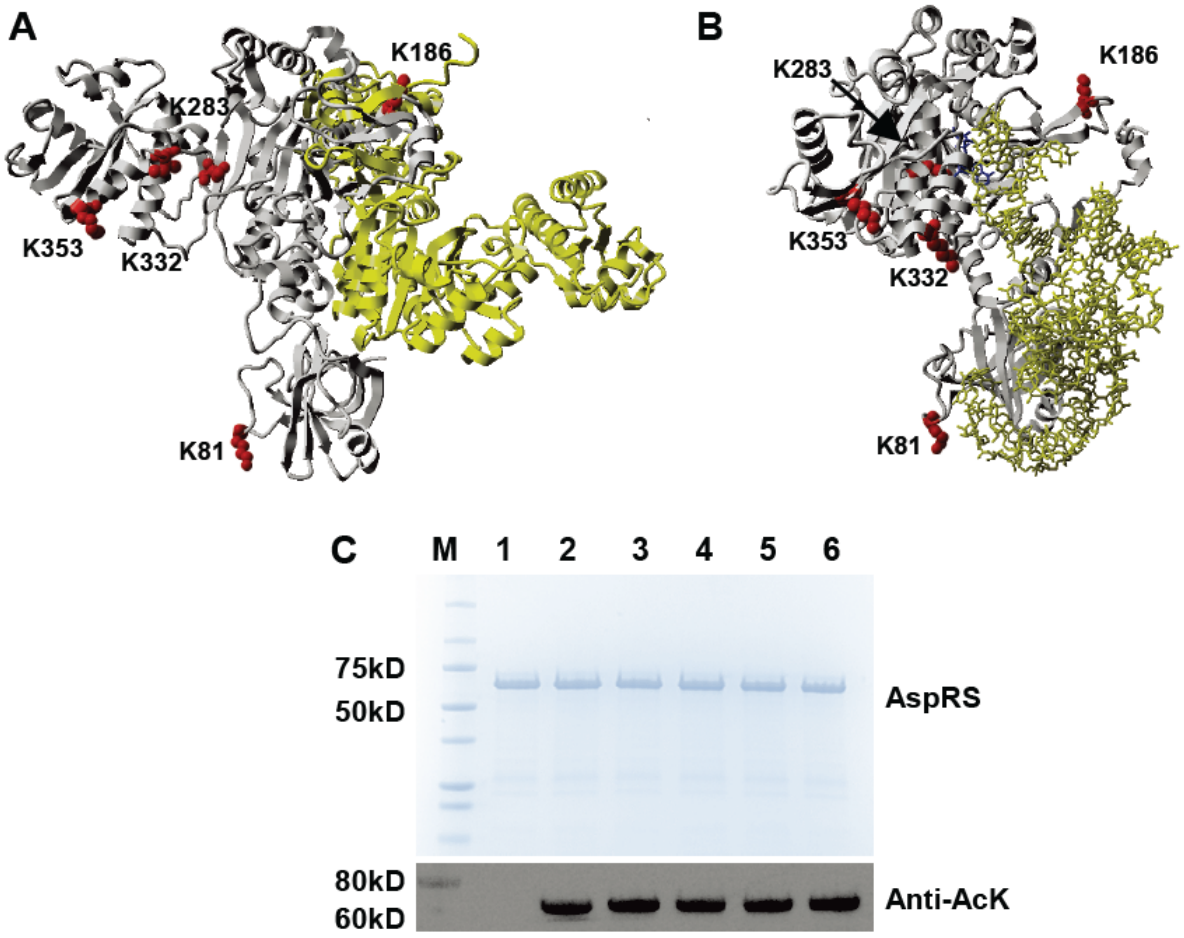


Figure S9 *E. coli* AspRS acetylation. A) Mapping of acetylated lysine residues on the crystal structure of the AspRS dimer (PDB ID: 1EQR). Two AspRS monomers are in yellow and grey, separately. Selected residues are in red. B) Mapping of acetylated lysine residues on the crystal structure of AspRS complexed with $tRNA^{Asp}$ (PDB ID: 1C0A). AspRS is in grey. $tRNA^{Asp}$ is in yellow. Aspartyl-adenylate is in blue. C) SDS-PAGE and western blotting analyses of purified AspRS and its variants. Lane 1, wild-type AspRS; lane 2, AspRS-81AcK; lane 3, AspRS-186AcK;lane 4, AspRS-283AcK; lane 5; AspRS-332AcK; lane 6, AspRS-353AcK. The same amounts of proteins were loaded.

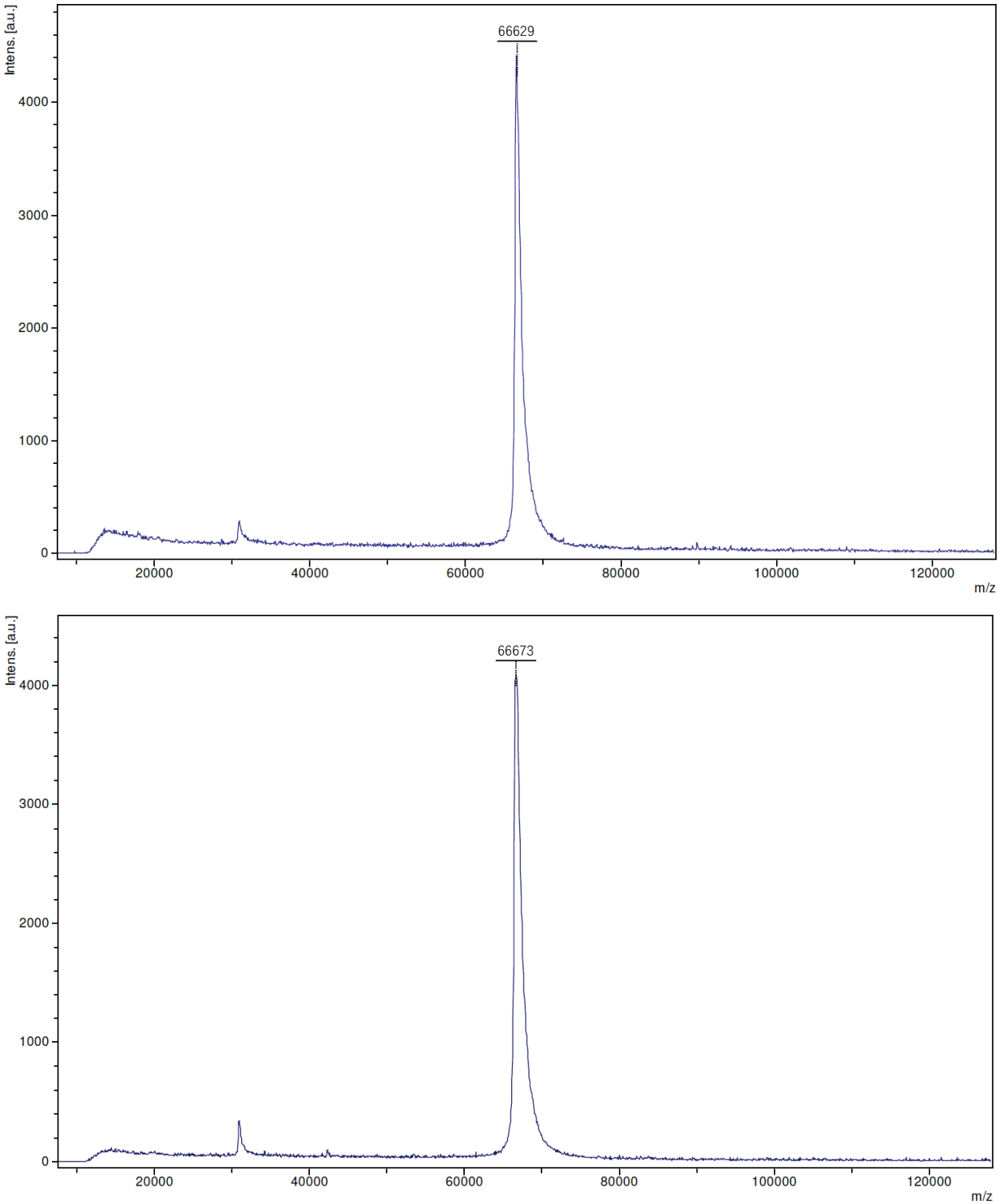
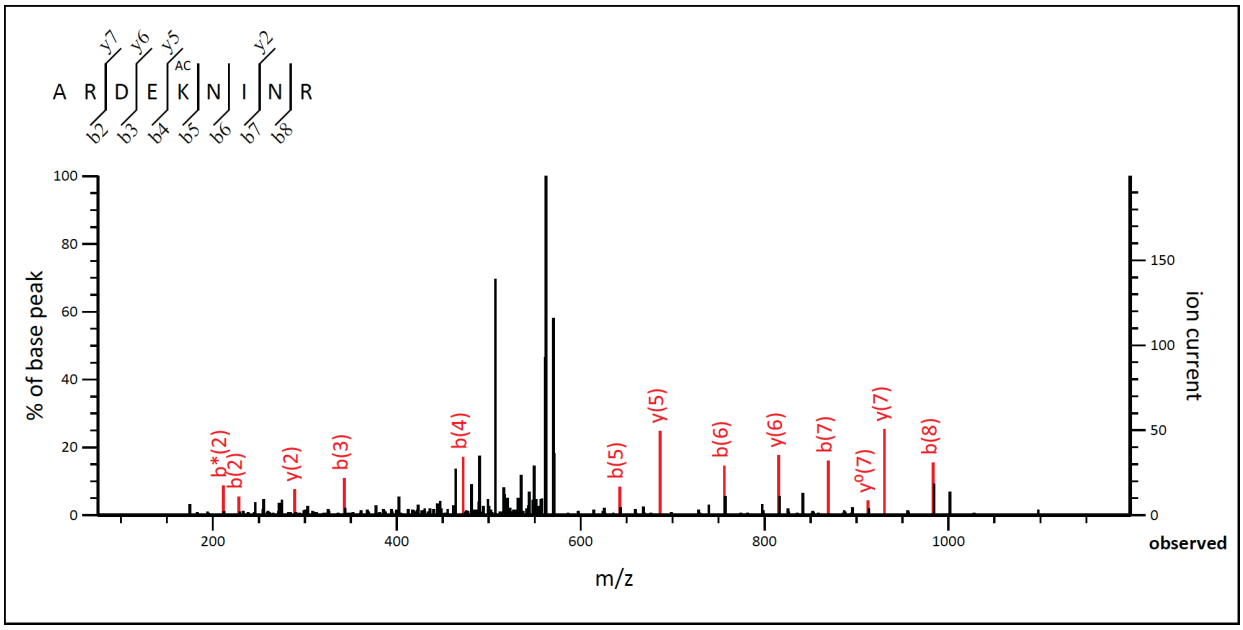
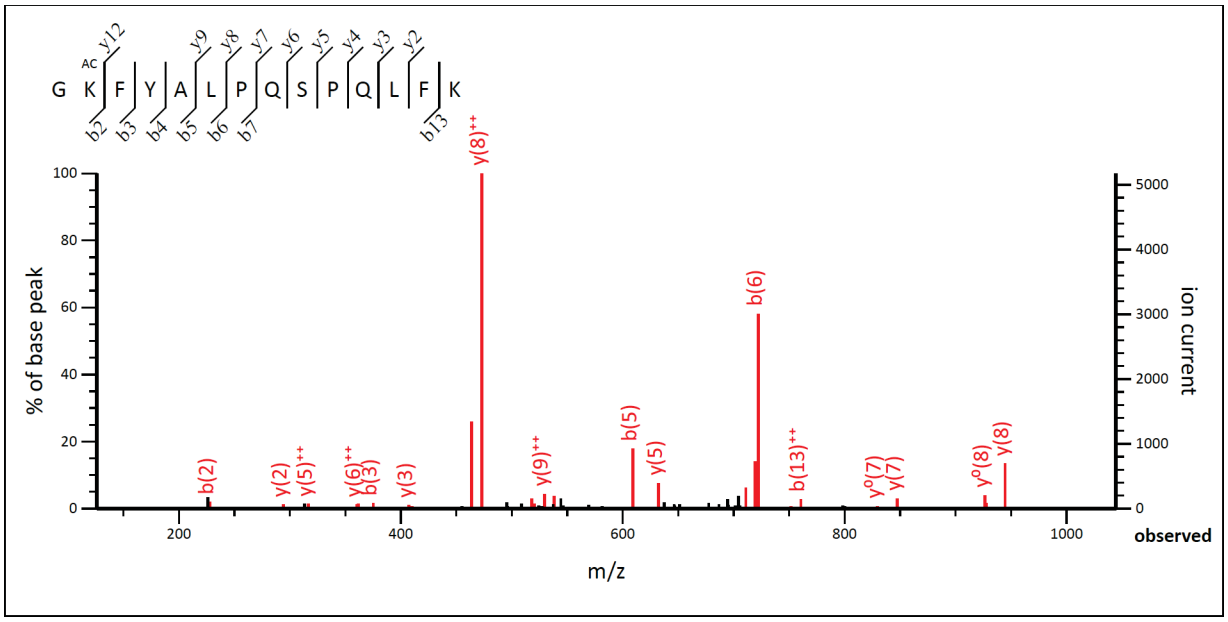


Figure S10 MALDI-TOF mass spectra of purified wild-type AspRS (upper, 66629 Da) and AspRS81-AcK (lower, 66673 Da) as a representative acetylated AspRS variant. The single peak of acetylated AspRS (ΔMW = 44 Da) indicated homogeneous incorporation of acetyllysine.



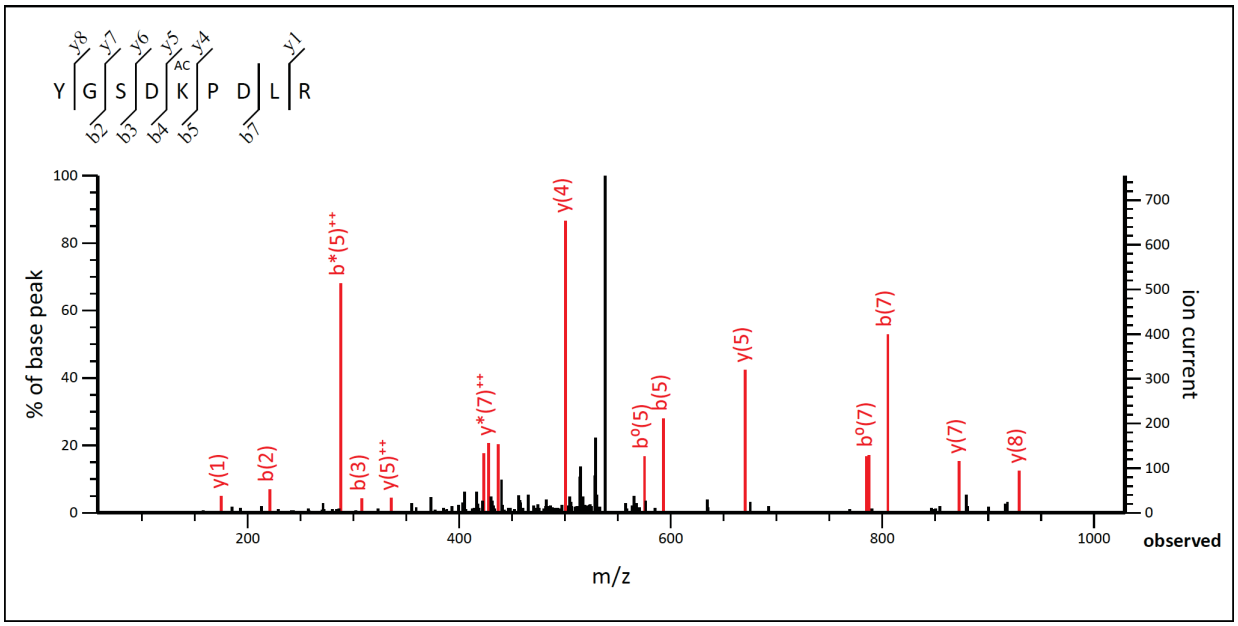
#	b	b^{++}	Seq.	y	y^{++}	#
1	72.0444	36.5258	A			9
2	*228.1455*	114.5764	R	1086.5650	543.7862	8
3	*343.1724*	172.0899	D	**930.4639**	465.7356	7
4	*472.2150*	236.6112	E	**815.4370**	408.2221	6
5	*642.3206*	321.6639	AcK	**686.3944**	343.7008	5
6	*756.3635*	378.6854	N	516.2889	258.6481	4
7	*869.4476*	435.2274	I	402.2459	201.6266	3
8	*983.4905*	492.2489	N	**289.1619**	145.0846	2
9			R	175.1190	88.0631	1

Figure S11 LC-MS/MS analysis of AspRS 81-AcK. The tandem mass spectrum of the peptide (residues 77-85) ARDEK^{AC}NINR from purified AspRS 81-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ionseries. Matched peaks are in red.



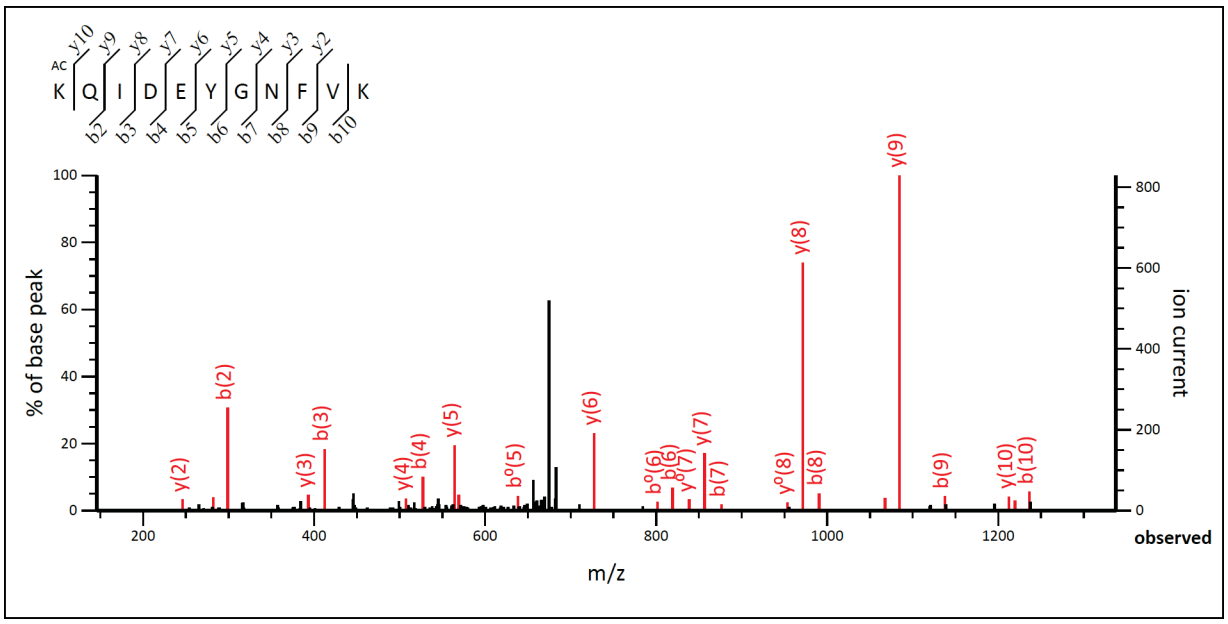
#	b	b^{++}	Seq.	y	y^{++}	#
1	58.0287	29.5180	G			14
2	*228.1343*	114.5708	AcK	1608.8784	804.9429	13
3	*375.2027*	188.1050	F	1438.7729	**719.8901**	12
4	*538.2660*	269.6366	Y	1291.7045	646.3559	11
5	*609.3031*	305.1552	A	1128.6412	564.8242	10
6	*722.3872*	**361.6972**	L	1057.6041	**529.3057**	9
7	819.4400	**410.2236**	P	*944.5200*	**472.7636**	8
8	947.4985	474.2529	Q	*847.4672*	424.2373	7
9	1034.5306	**517.7689**	S	*719.4087*	**360.2080**	6
10	1131.5833	566.2953	P	*632.3766*	**316.6920**	5
11	1259.6419	630.3246	Q	535.3239	268.1656	4
12	1372.7260	686.8666	L	*407.2653*	204.1363	3
13	1519.7944	**760.4008**	F	*294.1812*	147.5942	2
14			K	147.1128	74.0600	1

Figure S12 LC-MS/MS analysis of AspRS 186-AcK. The tandem mass spectrum of the peptide (residues 185-198) $GK^{AC}FYALPQSPQLFK$ from purified AspRS 186-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red.



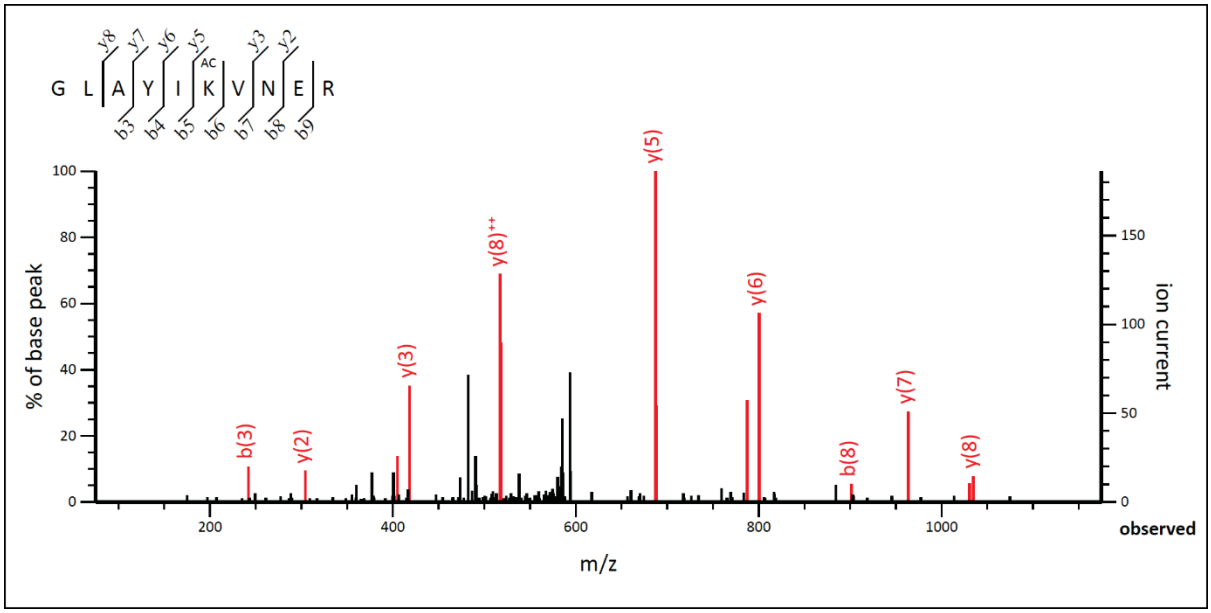
#	b	b^{++}	Seq.	y	y^{++}	#
1	164.0706	82.5389	Y			9
2	**221.0921**	111.0497	G	*929.4687*	465.2380	8
3	**308.1241**	154.5657	S	*872.4472*	**436.7272**	7
4	**423.1510**	212.0792	D	*785.4152*	393.2112	6
5	**593.2566**	297.1319	AcK	*670.3883*	**335.6978**	5
6	690.3093	345.6583	P	*500.2827*	250.6450	4
7	**805.3363**	403.1718	D	403.2300	202.1186	3
8	918.4203	459.7138	L	*288.2030*	144.6051	2
9			R	*175.1190*	88.0631	1

Figure S13 LC-MS/MS analysis of AspRS 283-AcK. The tandem mass spectrum of the peptide (residues 279-287) YGSDK^{AC}PDLR from purified AspRS 283-AcK. K^{AC} denotes AcK incorporation.The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red.



#	b	b^{++}	Seq.	y	y^{++}	#
1	171.1128	86.0600	AcK			11
2	*299.1714*	150.0893	Q	*1212.5895*	606.7984	10
3	*412.2554*	206.6314	I	*1084.5310*	542.7691	9
4	*527.2824*	264.1448	D	*971.4469*	486.2271	8
5	656.3250	328.6661	E	*856.4199*	428.7136	7
6	*819.3883*	410.1978	Y	*727.3774*	364.1923	6
7	*876.4098*	438.7085	G	*564.3140*	282.6606	5
8	*990.4527*	495.7300	N	*507.2926*	254.1499	4
9	*1137.5211*	569.2642	F	*393.2496*	197.1285	3
10	*1236.5895*	618.7984	V	*246.1812*	123.5942	2
11			K	147.1128	74.0600	1

Figure S14 LC-MS/MS analysis of AspRS 283-AcK. The tandem mass spectrum of the peptide (residues 332-342) K^{AC}QIDEYGNFVK from purified AspRS 332-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red.



#	b	b^{++}	Seq.	y	y^{++}	#
1	58.0287	29.5180	G			10
2	171.1128	86.0600	L	1147.6470	574.3271	9
3	*242.1499*	121.5786	A	*1034.5629*	517.7851	8
4	*405.2132*	203.1103	Y	*963.5258*	482.2665	7
5	*518.2973*	259.6523	I	*800.4625*	400.7349	6
6	*688.4028*	344.7051	AcK	*687.3784*	344.1928	5
7	*787.4713*	394.2393	V	*517.2729*	259.1401	4
8	*901.5142*	451.2607	N	*418.2045*	209.6059	3
9	*1030.5568*	515.7820	E	*304.1615*	152.5844	2
10			R	175.1190	88.0631	1

Figure S15 LC-MS/MS analysis of AspRS 353-AcK. The tandem mass spectrum of the peptide (residues 348-357) GLAYIK^{AC}VNER from purified AspRS 353-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red.

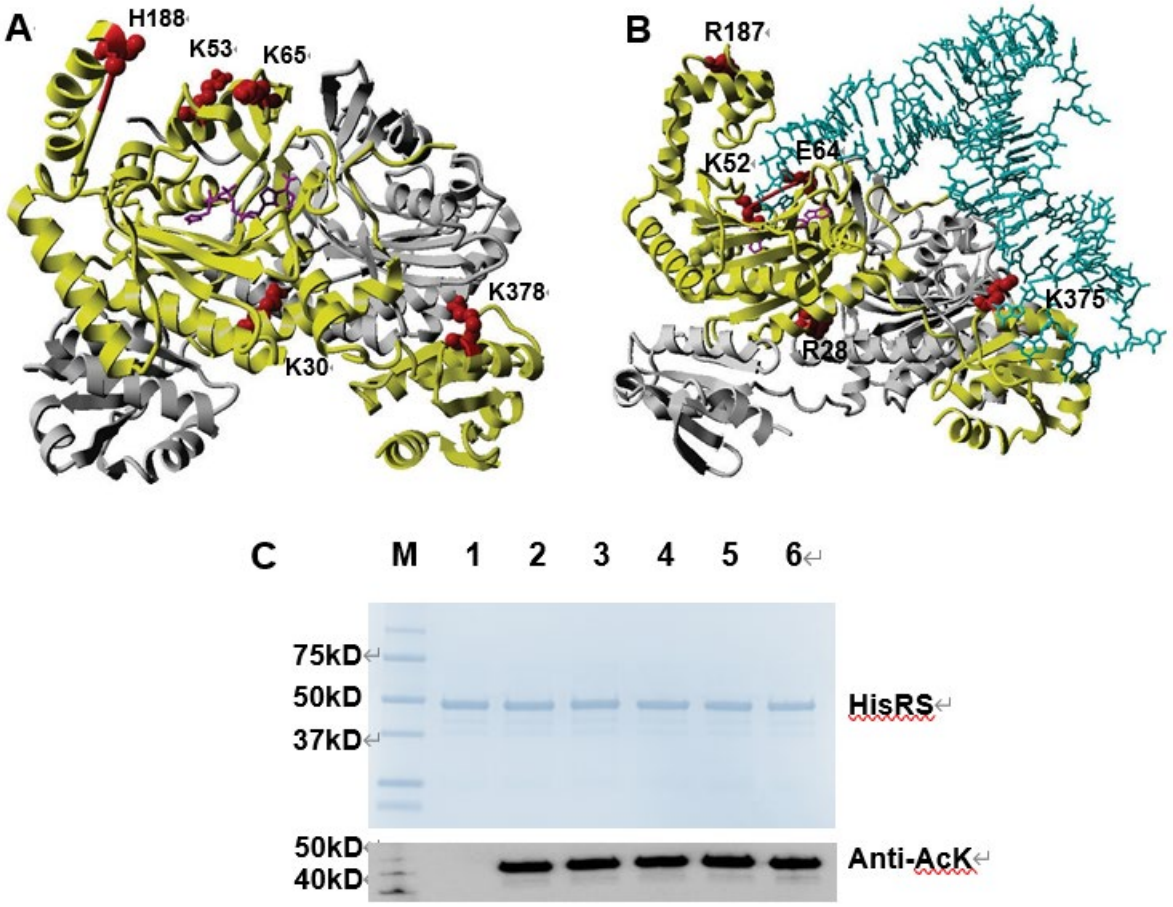


Figure S1 *E. coli* HisRS acetylation. A) Mapping of acetylated lysine residues on the crystal structure of E. coli HisRS (PDB ID: 1KMM). Two HisRS monomers are in yellow and grey, separately. Selected residues are in red. Histidyl-adenylate is in magenta. Residues from K189 to P223 were disordered, so the adjacent H188 was marked to show the approximate location of K189. B) Mapping of acetylated lysine residues on the crystal structure of *T. thermophilus* HisRScomplexed with tRNA (PDB ID: 4RDX). $tRNA^{His}$ is in cyan. The counterparts of selected residues(R28 for K30, K52 for K53, K63 for K65, R187 for K189, and K375 for K378) are in red. Histidineand adenylate are in magenta. Residues from G53 to K63 were disordered, so the adjacent E64was marked to show the approximate location of K63. C) SDS-PAGE and western blotting analyses of purified HisRS and its variants. Lane 1, wild-type HisRS; lane 2, HisRS-30AcK; lane3, HisRS-53AcK; lane 4, HisRS-65AcK; lane 5; HisRS-189AcK; lane 6, HisRS-378AcK. The sameamounts of proteins were loaded.

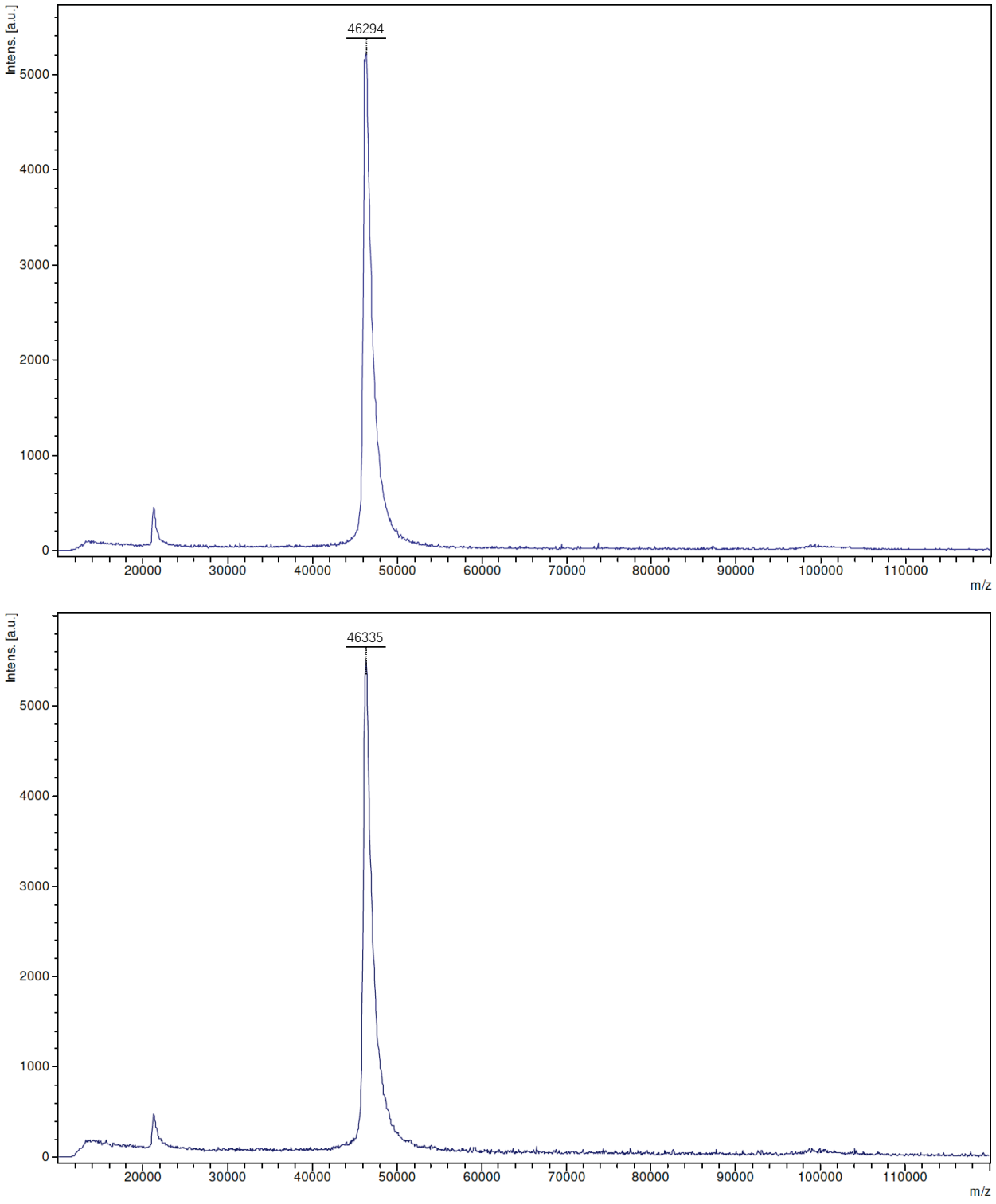
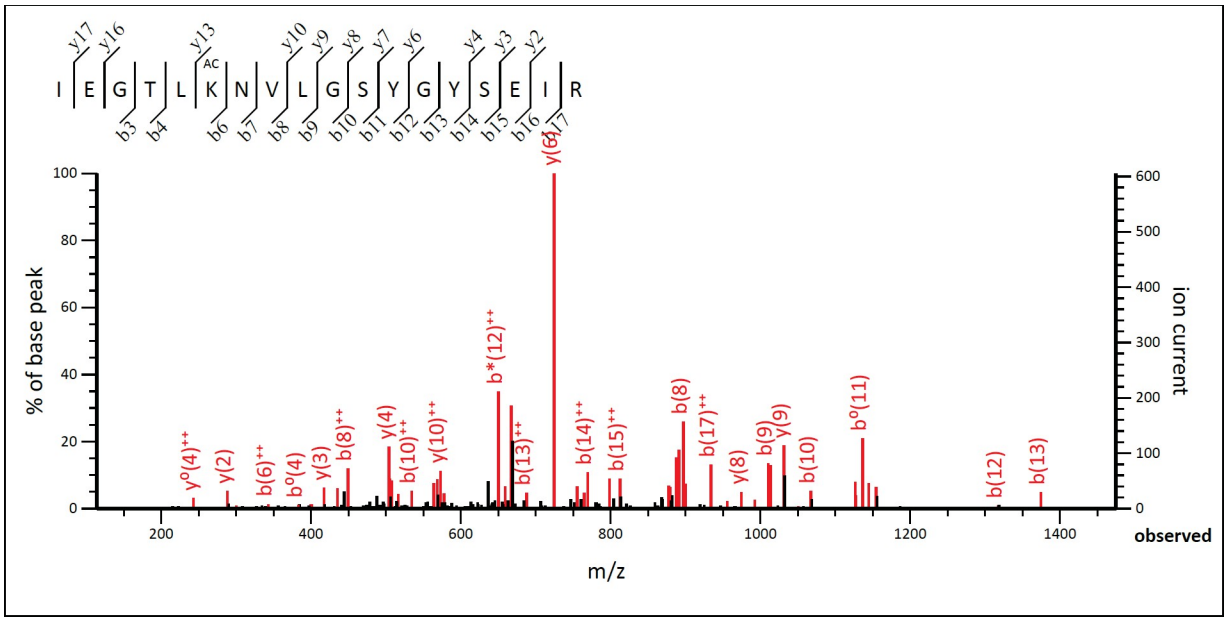
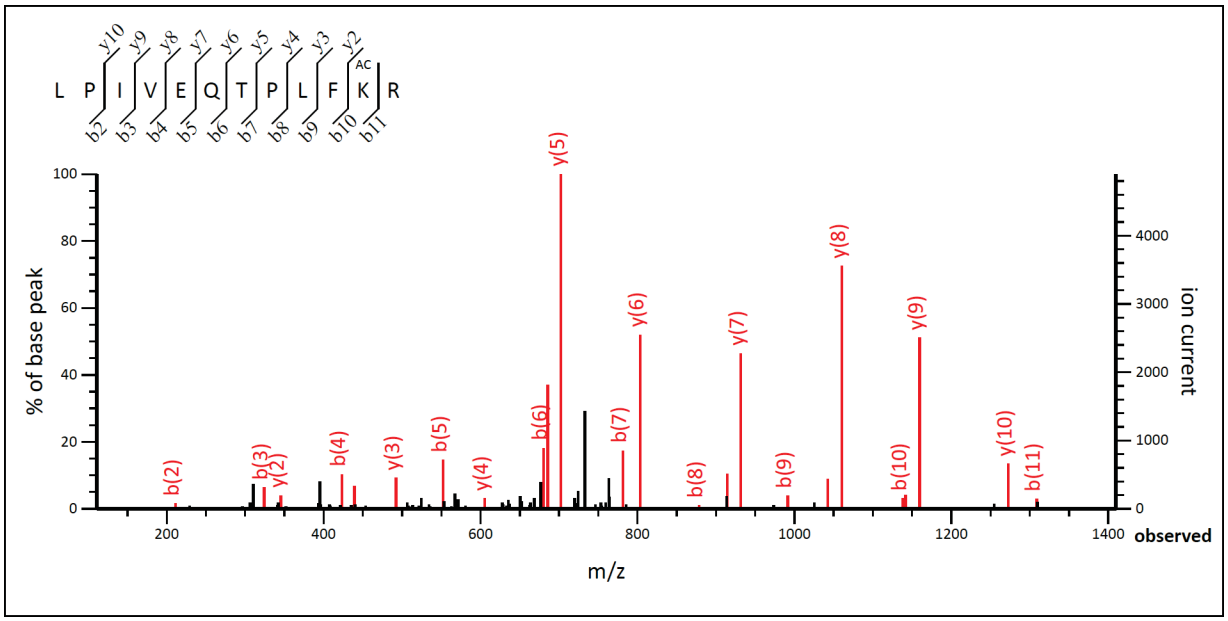


Figure S17 MALDI-TOF mass spectra of purified wild-type HisRS (upper, 46294 Da) and HisRS 30-AcK (lower, 46335 Da) as a representative acetylated HisRS variant. The single peak of acetylated HisRS (ΔMW = 41 Da) indicated homogeneous incorporation of acetyllysine.



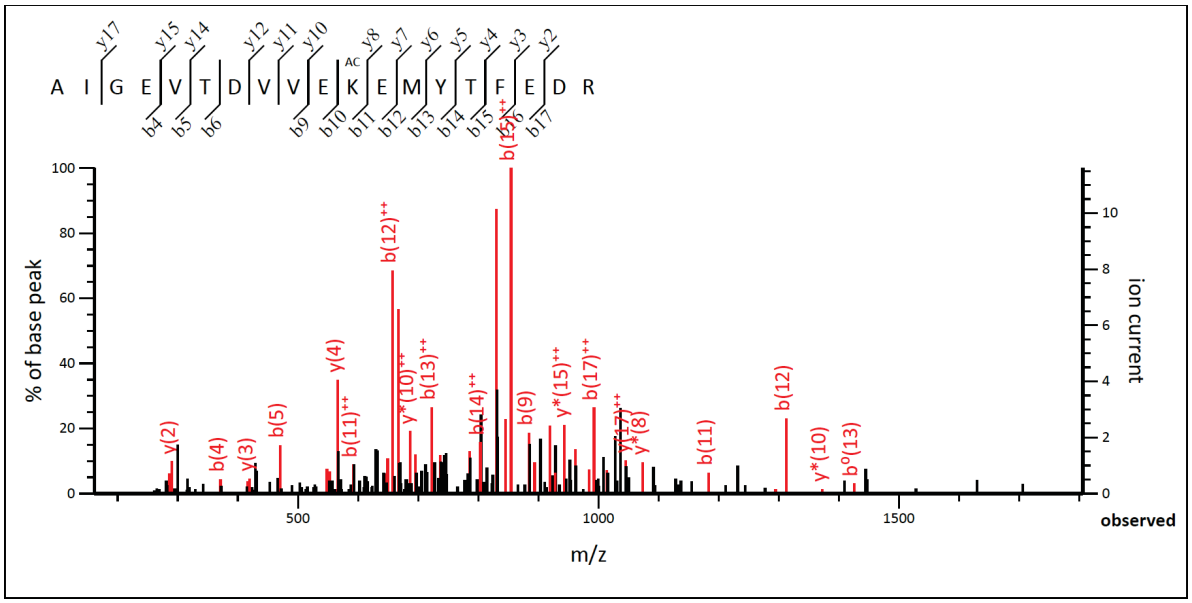
#	b	b^{++}	Seq.	y	y^{++}	#
1	114.0913	57.5493	I			18
2	*243.1339*	122.0706	E	1927.9760	964.4916	17
3	*300.1554*	150.5813	G	1798.9334	**899.9703**	16
4	*401.2031*	201.1052	T	1741.9119	871.4596	15
5	514.2871	257.6472	L	1640.8642	820.9358	14
6	684.3927	*342.7000*	AcK	1527.7802	**764.3937**	13
7	*798.4356*	*399.7214*	N	1357.6747	679.3410	12
8	*897.5040*	*449.2556*	V	1243.6317	622.3195	11
9	*1010.5881*	*505.7977*	L	**1144.5633**	**572.7853**	10
10	*1067.6095*	*534.3084*	G	**1031.4793**	**516.2433**	9
11	*1154.6416*	*577.8244*	S	**974.4578**	487.7325	8
12	*1317.7049*	*659.3561*	Y	**887.4258**	444.2165	7
13	*1374.7264*	*687.8668*	G	**724.3624**	362.6849	6
14	1537.7897	*769.3985*	Y	**667.3410**	334.1741	5
15	1624.8217	*812.9145*	S	**504.2776**	252.6425	4
16	1753.8643	*877.4358*	E	**417.2456**	209.1264	3
17	1866.9484	*933.9778*	I	**288.2030**	144.6051	2
18			R	175.1190	88.0631	1

Figure S18 LC-MS/MS analysis of HisRS 30-AcK. The tandem mass spectrum of the peptide (residues 25-42) IEGTLK^{AC}NVLGSYGYSEIR from purified HisRS 30-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red.



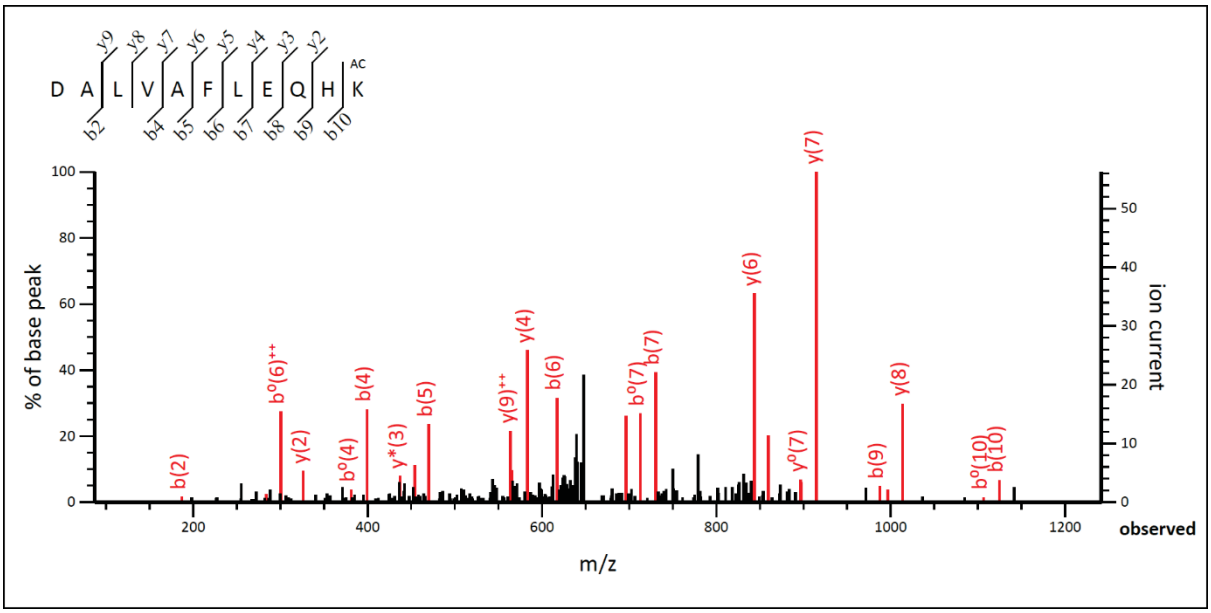
#	b	b^{++}	Seq.	y	y^{++}	#
1	114.0913	57.5493	L			12
2	*211.1441*	106.0757	P	1369.7838	685.3955	11
3	*324.2282*	162.6177	I	**1272.7311**	636.8692	10
4	*423.2966*	212.1519	V	**1159.6470**	580.3271	9
5	*552.3392*	276.6732	E	**1060.5786**	530.7929	8
6	*680.3978*	340.7025	Q	**931.5360**	466.2716	7
7	*781.4454*	391.2264	T	**803.4774**	402.2423	6
8	*878.4982*	439.7527	P	**702.4297**	351.7185	5
9	*991.5823*	496.2948	L	**605.3770**	303.1921	4
10	*1138.6507*	569.8290	F	**492.2929**	246.6501	3
11	*1308.7562*	654.8817	AcK	**345.2245**	173.1159	2
12			R	175.1190	88.0631	1

Figure S19 LC-MS/MS analysis of HisRS 53-AcK. The tandem mass spectrum of the peptide (residues 43-54) LPIVEQTPLFK^{AC}R from purified HisRS 53-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red.



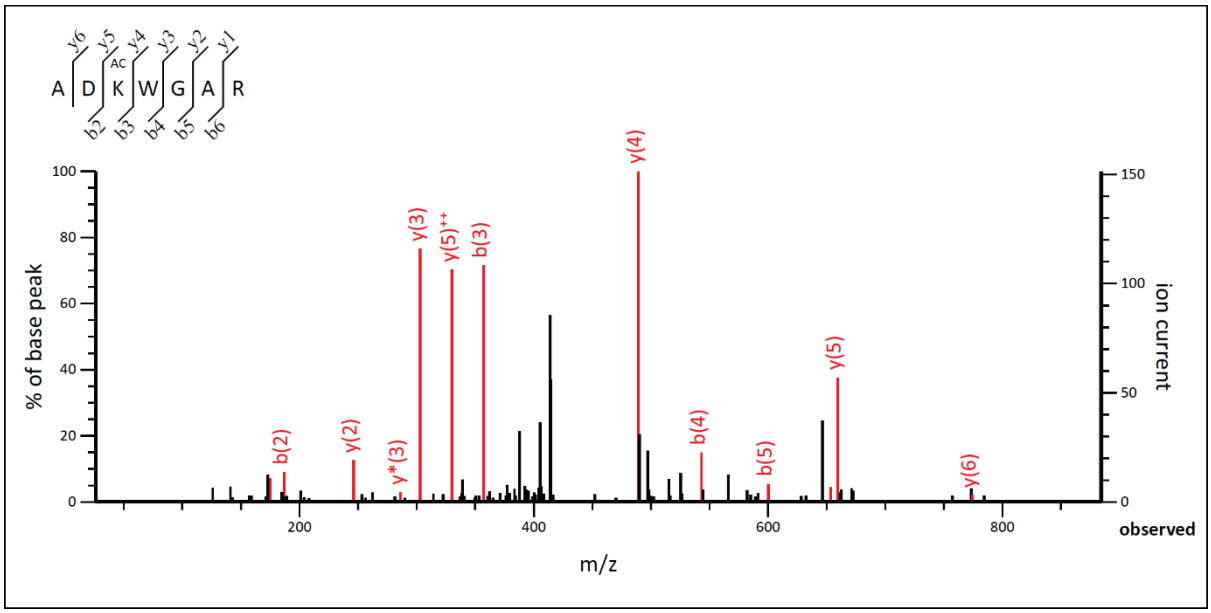
#	b	b^{++}	Seq.	y	y^{++}	#
1	72.0444	36.5258	A			19
2	185.1285	93.0679	I	2202.0271	1101.5172	18
3	242.1499	121.5786	G	2088.9430	1044.9752	17
4	*371.1925*	186.0999	E	2031.9216	1016.4644	16
5	*470.2609*	235.6341	V	1902.8790	951.9431	15
6	571.3086	*286.1579*	T	1803.8106	902.4089	14
7	*686.3355*	343.6714	D	1702.7629	851.8851	13
8	*785.4040*	393.2056	V	1587.7359	794.3716	12
9	*884.4724*	442.7398	V	1488.6675	744.8374	11
10	*1013.5150*	507.2611	E	1389.5991	695.3032	10
11	*1183.6205*	*592.3139*	AcK	1260.5565	630.7819	9
12	*1312.6631*	*656.8352*	E	1090.4510	545.7291	8
13	1443.7036	*722.3554*	M	**961.4084**	481.2078	7
14	1606.7669	*803.8871*	Y	**830.3679**	415.6876	6
15	1707.8146	*854.4109*	T	**667.3046**	334.1559	5
16	1854.8830	*927.9451*	F	**566.2569**	283.6321	4
17	1983.9256	*992.4664*	E	**419.1885**	210.0979	3
18	2098.9525	1049.9799	D	**290.1459**	145.5766	2
19			R	175.1190	88.0631	1

Figure S20 LC-MS/MS analysis of HisRS 65-AcK. The tandem mass spectrum of the peptide (residues 55-73) AIGEVTDVVEK^{AC}EMYTFEDR from purified HisRS 65-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red.



#	b	b^{++}	Seq.	y	y^{++}	#
1	116.0342	58.5207	D			11
2	*187.0713*	94.0393	A	1197.6626	599.3350	10
3	*300.1554*	150.5813	L	1126.6255	563.8164	9
4	*399.2238*	200.1155	V	**1013.5415**	507.2744	8
5	*470.2609*	235.6341	A	**914.4730**	457.7402	7
6	*617.3293*	309.1683	F	**843.4359**	422.2216	6
7	*730.4134*	365.7103	L	**696.3675**	348.6874	5
8	*859.4560*	430.2316	E	**583.2835**	292.1454	4
9	*987.5146*	494.2609	Q	**454.2409**	227.6241	3
10	*1124.5735*	562.7904	H	**326.1823**	163.5948	2
11			AcK	189.1234	95.0653	1

Figure S21 LC-MS/MS analysis of HisRS 189-AcK. The tandem mass spectrum of the peptide (residues 179-189) DALVAFLEQHK^{AC} from purified HisRS 189-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red.



#	b	b++	Seq.	y	y++	#
1	72.0444	36.5258	A			7
2	**187.0713**	94.0393	D	***774.3893***	387.6983	6
3	**357.1769**	179.0921	AcK	***659.3624***	330.1848	5
4	**543.2562**	272.1317	W	***489.2568***	245.1321	4
5	**600.2776**	300.6425	G	***303.1775***	152.0924	3
6	671.3148	336.1610	A	***246.1561***	123.5817	2
7			R	***175.1190***	88.0631	1

Figure S22 LC-MS/MS analysis of HisRS 378-AcK. The tandem mass spectrum of the peptide (residues 376-382) ADK^{AC}WGAR from purified HisRS 378-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red.

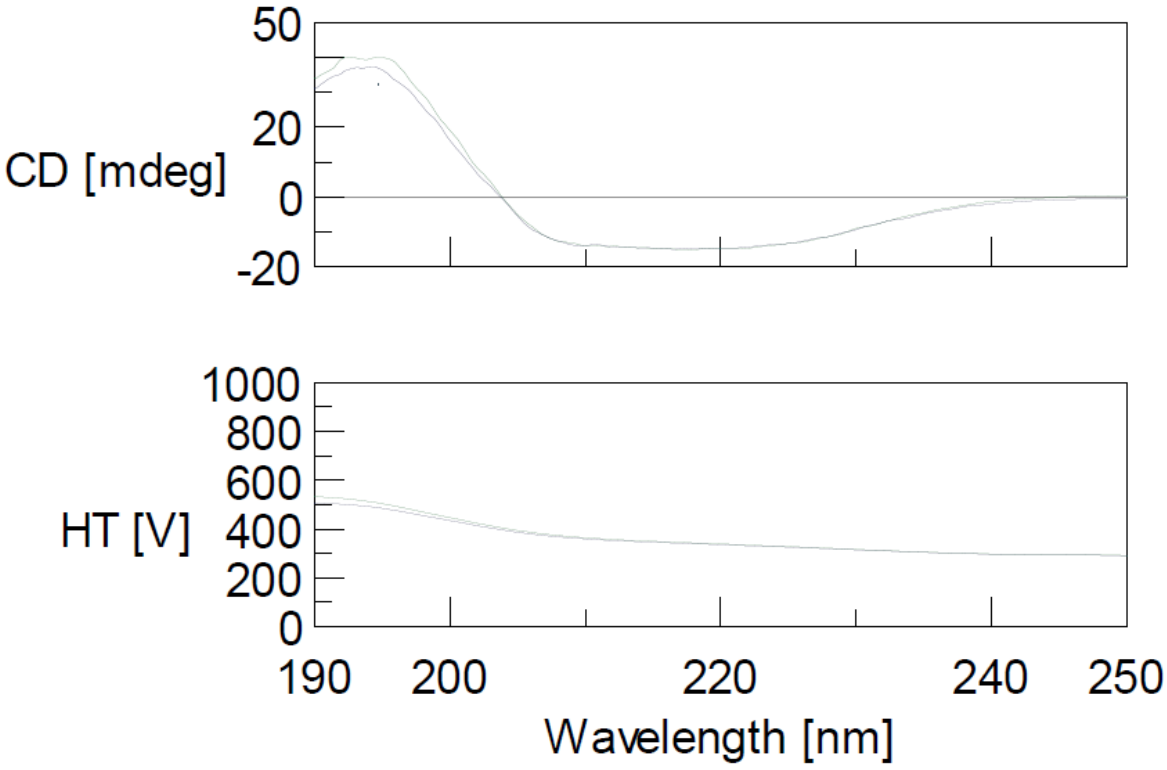


Figure S23 The circular dichroism (CD) spectra of the wild-type HisRS (blue lines) and its 378AcK variant (green lines). The CD spectra were recorded on a J-1500 CD Spectrometer. Purified enzymes were diluted to a concentration of 0.1 mg/ml in 5 mM Tris-HCl pH 7.8, 0.1 M KCl, and scanned from 190 nm to 250 nm with a 20 nm/min speed. Scanning was performed three times for each sample and the average was plotted.

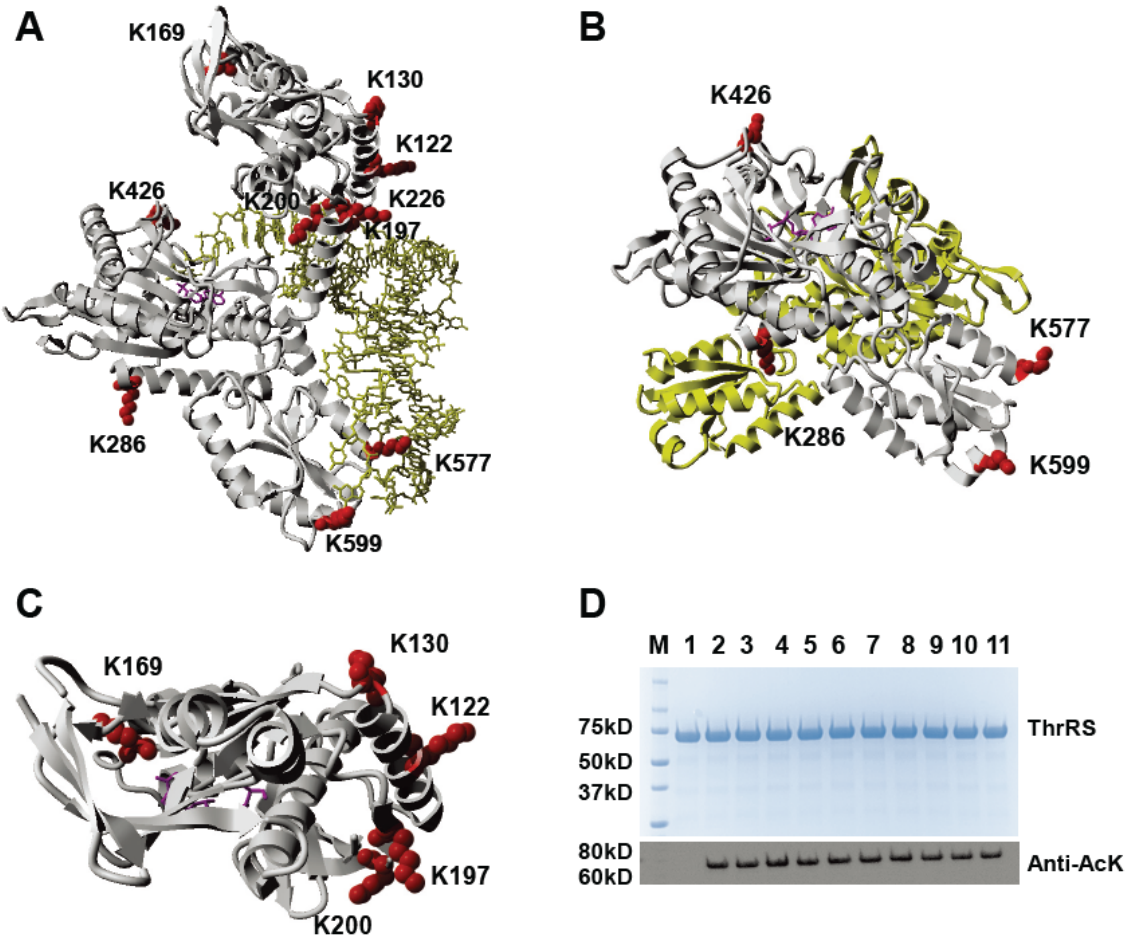


Figure S24 *E. coli* ThrRS acetylation. A) Mapping of acetylated lysine residues on the crystal structure of *E. coli* ThrRS complexed with $tRNA^{Thr}$ (PDB ID: 1QF6). ThrRS is in grey, and $tRNA^{Thr}$is in yellow. Selected lysine residues are in red. Adenylate is in magenta. B) Mapping of acetylatedlysine residues on the crystal structure of the ThrRS dimer (PDB ID: 1EVL). Two truncated ThrRSmonomers are in yellow and grey, separately. The threonyl-adenylate analog is in magenta. C) Mapping of acetylated lysine residues in the N-terminal editing domain of ThrRS (PDB ID: 1TKG).The seryl-adenylated analog is in magenta. D) SDS-PAGE and western blotting analyses of purified ThrRS and its variants. Lane 1, wild-type ThrRS; lane 2, ThrRS-122AcK; lane 3, ThrRS-130AcK; lane 4, ThrRS-169AcK; lane 5; ThrRS-197AcK; lane 6, ThrRS-200AcK; lane 7, ThrRS-226AcK; lane 8, ThrRS-286AcK; lane 9, ThrRS-426AcK; lane 10, ThrRS-577AcK; lane 11, ThrRS-599AcK; The same amounts of proteins were loaded.

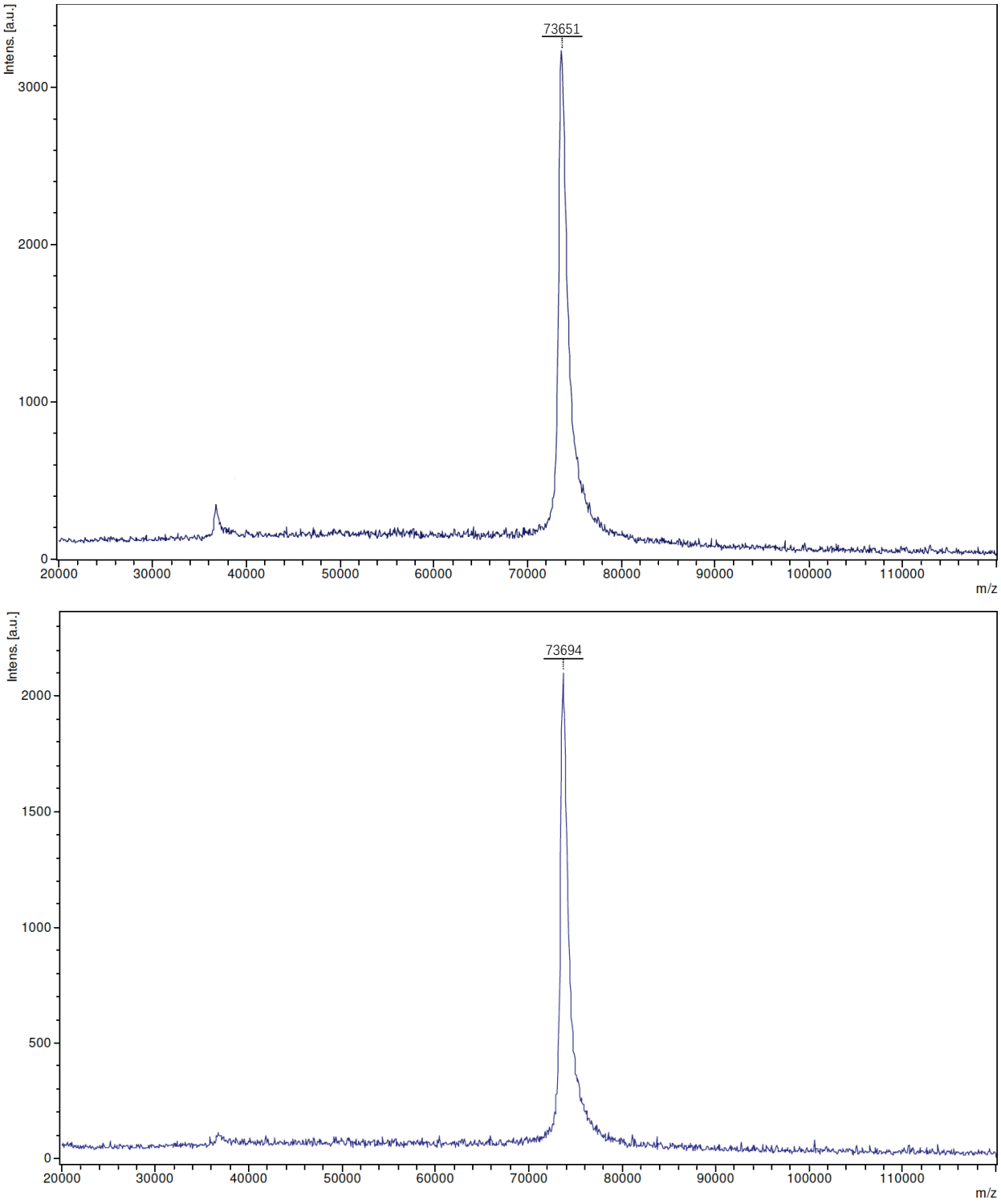
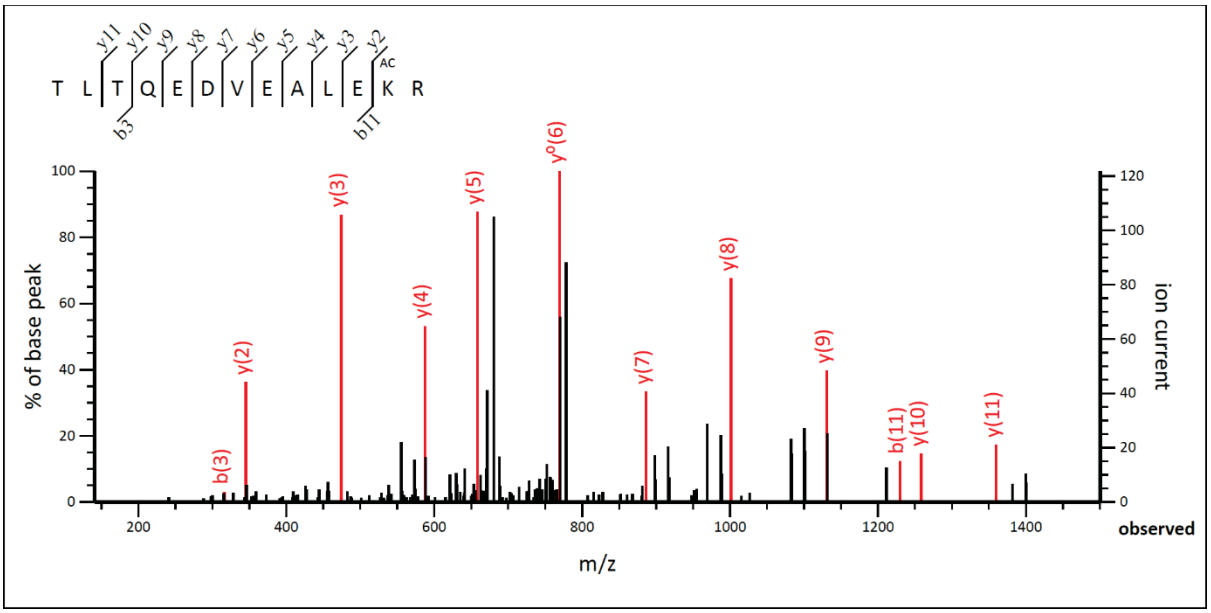
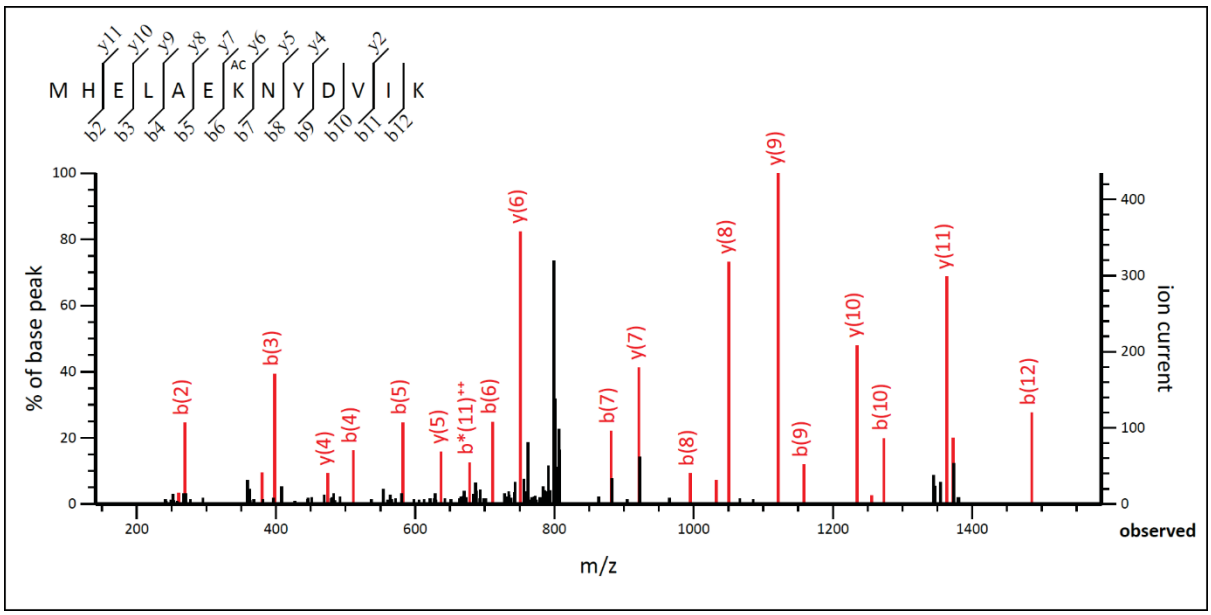


Figure S25 MALDI-TOF mass spectra of purified wild-type ThrRS (upper, 73651 Da) and ThrRS122-AcK (lower, 73694 Da) as a representative acetylated ThrRS variant. The single peak of acetylated ThrRS (ΔMW = 43 Da) indicated homogeneous incorporation of acetyllysine.



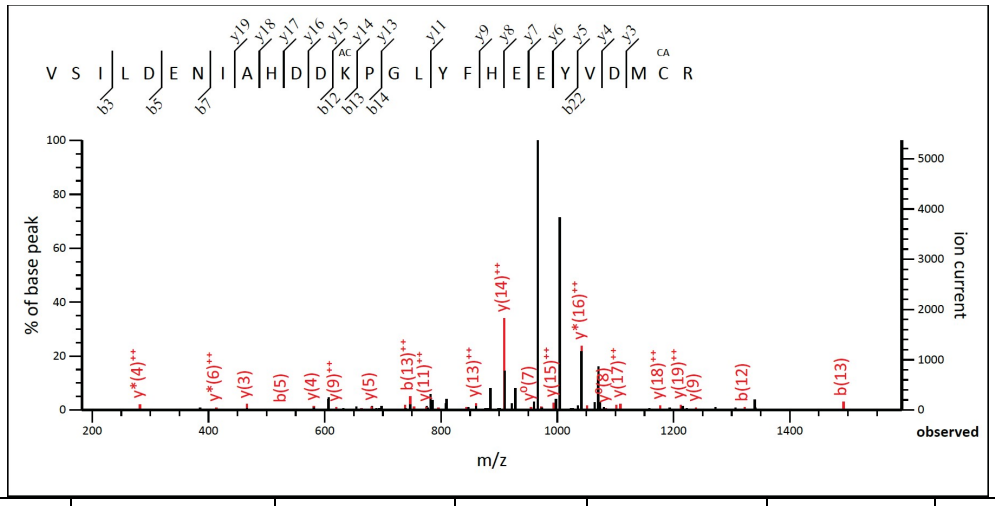
#	b	b^{++}	Seq.	y	y^{++}	#
1	102.0550	51.5311	T			13
2	215.1390	108.0731	L	1472.7591	736.8832	12
3	**316.1867**	158.5970	T	*1359.6751*	680.3412	11
4	444.2453	222.6263	Q	*1258.6274*	629.8173	10
5	573.2879	287.1476	E	*1130.5688*	565.7880	9
6	688.3148	344.6610	D	*1001.5262*	501.2667	8
7	787.3832	394.1953	V	*886.4993*	443.7533	7
8	916.4258	458.7165	E	787.4308	394.2191	6
9	987.4629	494.2351	A	*658.3883*	329.6978	5
10	1100.5470	550.7771	L	*587.3511*	294.1792	4
11	**1229.5896**	615.2984	E	*474.2671*	237.6372	3
12	1399.6951	700.3512	AcK	*345.2245*	173.1159	2
13			R	175.1190	88.0631	1

Figure S26 LC-MS/MS analysis of ThrRS 122-AcK. The tandem mass spectrum of the peptide (residues 111-123) TLTQEDVEALEK^{AC}R from purified ThrRS 122-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red.



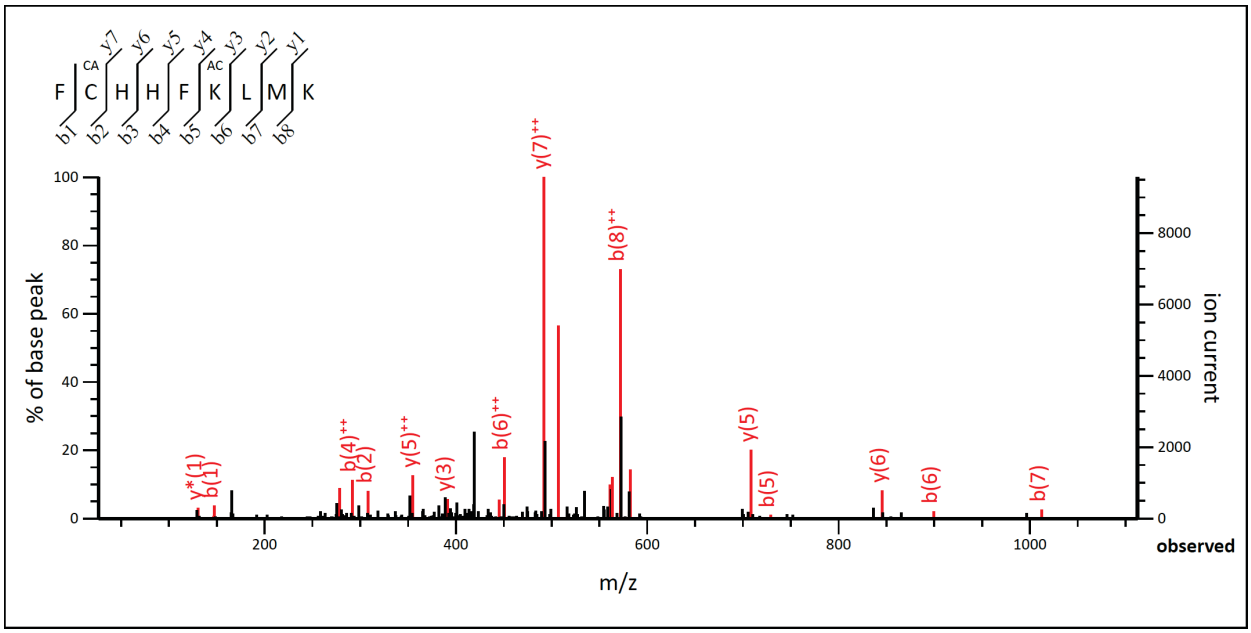
#	b	b^{++}	Seq.	y	y^{++}	#
1	132.0478	66.5275	M			13
2	*269.1067*	135.0570	H	1500.7693	750.8883	12
3	*398.1493*	199.5783	E	**1363.7104**	682.3588	11
4	*511.2333*	256.1203	L	**1234.6678**	617.8375	10
5	*582.2704*	291.6389	A	**1121.5837**	561.2955	9
6	*711.3130*	356.1602	E	**1050.5466**	525.7769	8
7	*881.4186*	441.2129	AcK	**921.5040**	461.2556	7
8	*995.4615*	498.2344	N	**751.3985**	376.2029	6
9	*1158.5248*	579.7660	Y	**637.3556**	319.1814	5
10	*1273.5518*	637.2795	D	**474.2922**	237.6498	4
11	*1372.6202*	686.8137	V	359.2653	180.1363	3
12	*1485.7042*	743.3558	I	**260.1969**	130.6021	2
13			K	147.1128	74.0600	1

Figure S27 LC-MS/MS analysis of ThrRS 130-AcK. The tandem mass spectrum of the peptide (residues 124-136) MHELAEK^{AC}NYDVIK from purified ThrRS 130-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red.



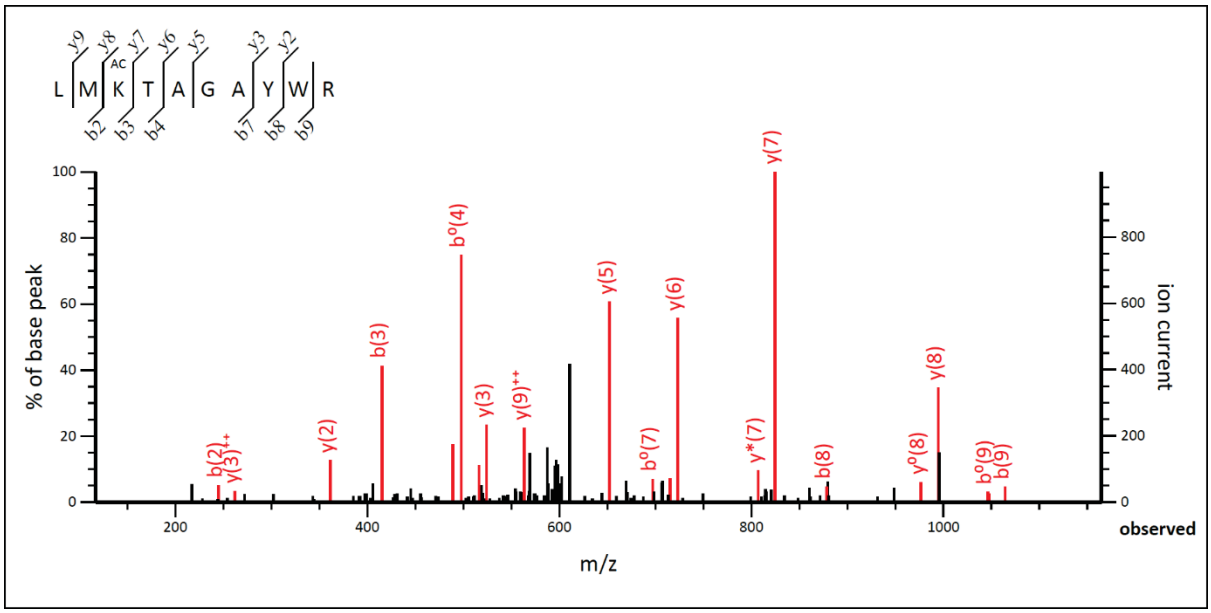
#	b	b^{++}	Seq.	y	y^{++}	#
1	100.0757	50.5415	V			27
2	187.1077	94.0575	S	3208.4350	1604.7212	26
3	*300.1918*	150.5995	I	3121.4030	1561.2051	25
4	*413.2758*	207.1416	L	3008.3189	1504.6631	24
5	*528.3028*	264.6550	D	2895.2349	1448.1211	23
6	657.3454	329.1763	E	2780.2079	1390.6076	22
7	771.3883	386.1978	N	2651.1653	1326.0863	21
8	884.4724	442.7398	I	2537.1224	1269.0648	20
9	955.5095	478.2584	A	2424.0383	*1212.5228*	19
10	1092.5684	546.7878	H	2353.0012	*1177.0043*	18
11	1207.5953	604.3013	D	2215.9423	*1108.4748*	17
12	*1322.6223*	661.8148	D	2100.9154	*1050.9613*	16
13	*1492.7278*	746.8675	AcK	1985.8884	*993.4479*	15
14	1589.7806	795.3939	P	1815.7829	*908.3951*	14
15	1646.8020	823.9047	G	1718.7301	*859.8687*	13
16	1759.8861	880.4467	L	1661.7087	831.3580	12
17	1922.9494	961.9784	Y	1548.6246	*774.8159*	11
18	2070.0178	1035.5126	F	1385.5613	693.2843	10
19	2207.0768	1104.0420	H	*1238.4929*	*619.7501*	9
20	2336.1194	1168.5633	E	*1101.4340*	551.2206	8
21	2465.1619	1233.0846	E	*972.3914*	486.6993	7
22	2628.2253	1314.6163	Y	*843.3488*	422.1780	6
23	2727.2937	1364.1505	V	*680.2854*	340.6464	5
24	2842.3206	1421.6640	D	*581.2170*	291.1122	4
25	2973.3611	1487.1842	M	*466.1901*	233.5987	3
26	3133.3918	1567.1995	C	335.1496	168.0784	2
27			R	175.1190	88.0631	1

Figure S28 LC-MS/MS analysis of ThrRS 169-AcK. The tandem mass spectrum of the peptide (residues 157-183) VSILDENIAHDDK^{AC}PGLYFHEEYVDMCR from purified ThrRS 169-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red



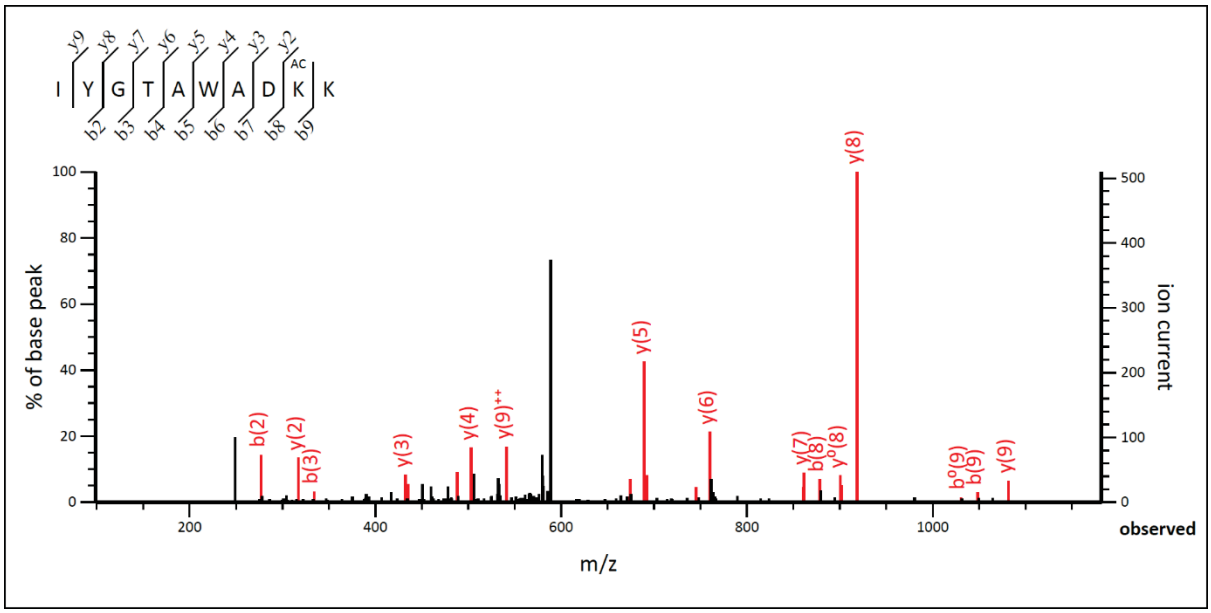
#	b	b++	Seq.	y	y++	#
1	*148.0757*	74.5415	F			9
2	*308.1063*	154.5568	C	1142.5598	571.7835	8
3	*445.1653*	223.0863	H	982.5291	491.7682	7
4	*582.2242*	291.6157	H	845.4702	423.2387	6
5	*729.2926*	365.1499	F	708.4113	354.7093	5
6	*899.3981*	450.2027	AcK	561.3429	281.1751	4
7	*1012.4822*	506.7447	L	391.2374	196.1223	3
8	1143.5227	572.2650	M	278.1533	139.5803	2
9			K	147.1128	147.1128	1

Figure S29 LC-MS/MS analysis of ThrRS 197-AcK. The tandem mass spectrum of the peptide (residues 192-200) FCHHFK^{AC}LMK from purified ThrRS 197-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red.



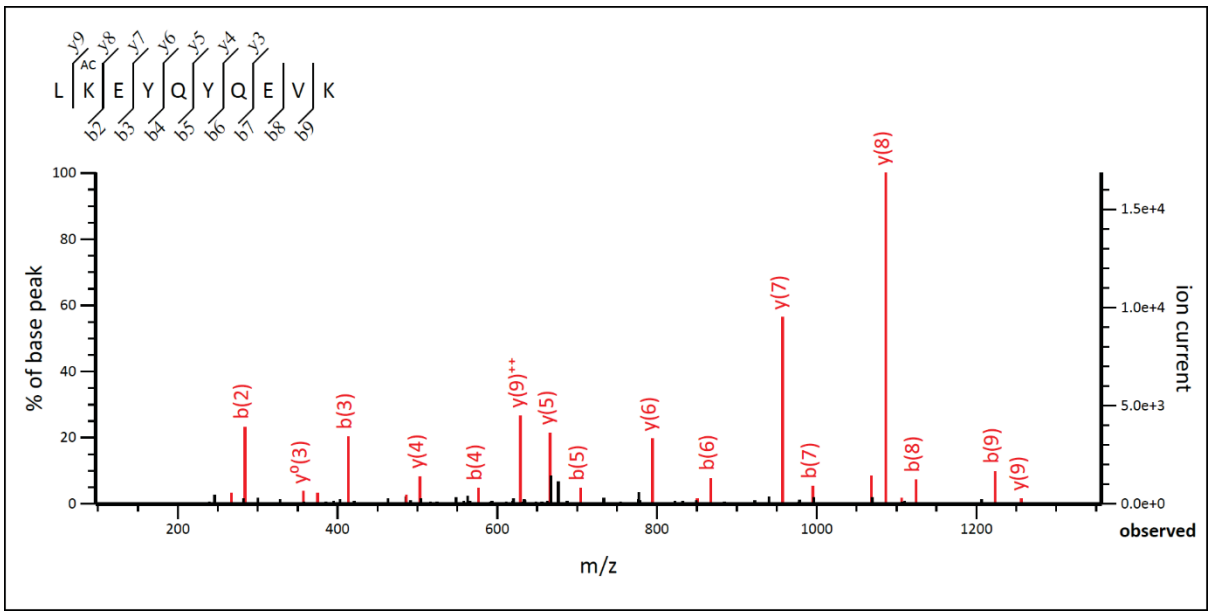
#	b	b^{++}	Seq.	y	y^{++}	#
1	114.0913	57.5493	L			10
2	*245.1318*	123.0696	M	1125.5510	**563.2791**	9
3	*415.2374*	208.1223	AcK	*994.5105*	**497.7589**	8
4	*516.2850*	258.6462	T	*824.4050*	412.7061	7
5	587.3221	294.1647	A	*723.3573*	362.1823	6
6	644.3436	322.6754	G	*652.3202*	326.6637	5
7	*715.3807*	358.1940	A	595.2987	298.1530	4
8	*878.4441*	439.7257	Y	*524.2616*	**262.6344**	3
9	*1064.5234*	532.7653	W	*361.1983*	181.1028	2
10			R	175.1190	88.0631	1

Figure S30 LC-MS/MS analysis of ThrRS 200-AcK. The tandem mass spectrum of the peptide (residues 198-207) LMK^{AC}TAGAYWR from purified ThrRS 200-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red.



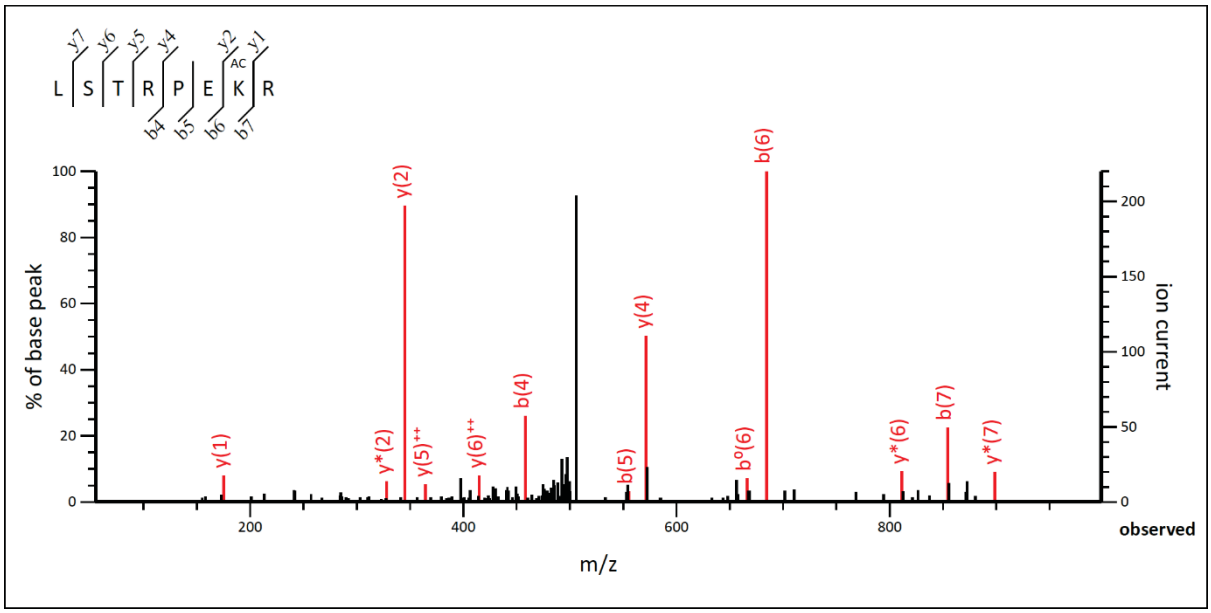
#	b	b^{++}	Seq.	y	y^{++}	#
1	114.0913	57.5493	I			10
2	277.1547	139.0810	Y	*1081.5313*	541.2693	9
3	334.1761	167.5917	G	*918.4680*	459.7376	8
4	435.2238	218.1155	T	*861.4465*	431.2269	7
5	506.2609	253.6341	A	*760.3988*	380.7030	6
6	692.3402	346.6738	W	*689.3617*	345.1845	5
7	763.3774	382.1923	A	*503.2824*	252.1448	4
8	878.4043	439.7058	D	*432.2453*	216.6263	3
9	1048.5098	524.7586	AcK	*317.2183*	159.1128	2
10			K	147.1128	74.0600	1

Figure S31 LC-MS/MS analysis of ThrRS 226-AcK. The tandem mass spectrum of the peptide (residues 218-227) IYGTAWADK^{AC}K from purified ThrRS 226-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red.



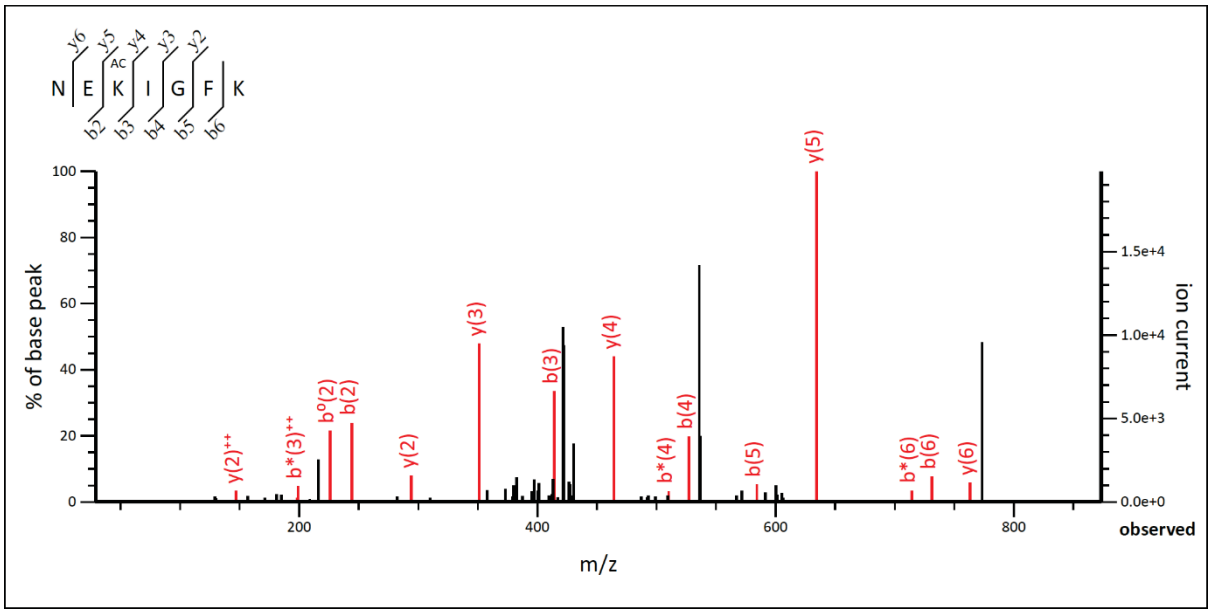
#	b	b^{++}	Seq.	y	y^{++}	#
1	114.0913	57.5493	L			10
2	*284.1969*	142.6021	AcK	*1256.6157*	628.8115	9
3	*413.2395*	207.1234	E	*1086.5102*	543.7587	8
4	*576.3028*	288.6550	Y	*957.4676*	479.2375	7
5	*704.3614*	352.6843	Q	*794.4043*	397.7058	6
6	*867.4247*	434.2160	Y	*666.3457*	333.6765	5
7	*995.4833*	498.2453	Q	*503.2824*	252.1448	4
8	*1124.5259*	562.7666	E	*375.2238*	188.1155	3
9	*1223.5943*	612.3008	V	246.1812	123.5942	2
10			K	147.1128	74.0600	1

Figure S32 LC-MS/MS analysis of ThrRS 286-AcK. The tandem mass spectrum of the peptide (residues 285-294) LK^{AC}EYQYQEVK from purified ThrRS 286-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red.



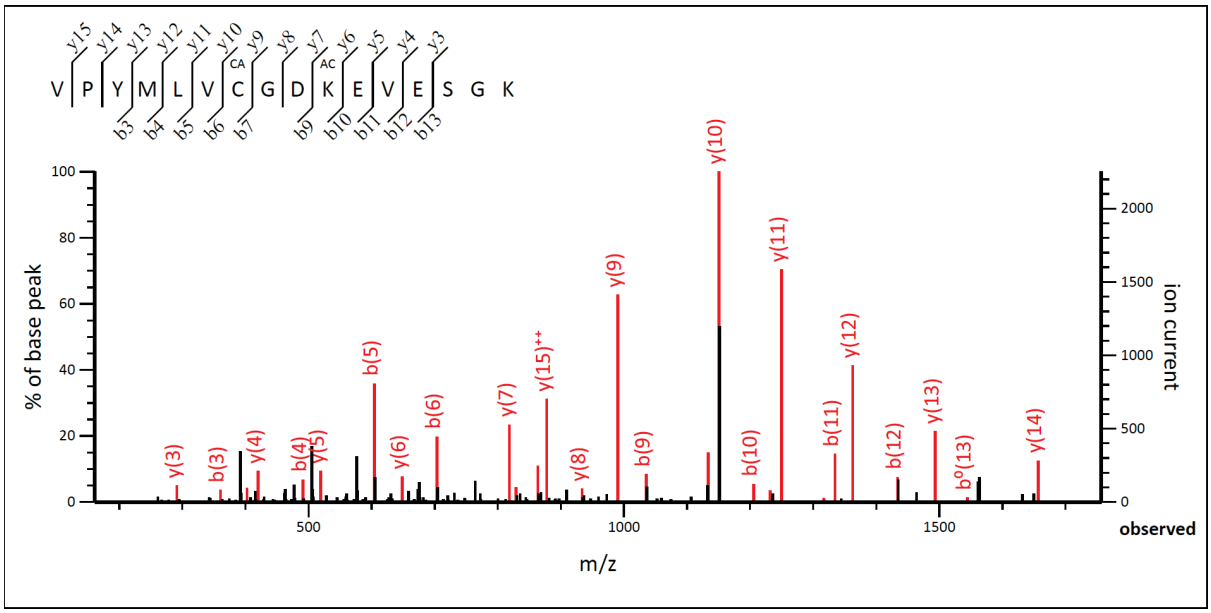
#	b	b++	Seq.	y	y++	#
1	114.0913	57.5493	L			8
2	201.1234	101.0653	S	915.5007	*458.2540*	7
3	302.1710	151.5892	T	828.4686	*414.7380*	6
4	*458.2722*	229.6397	R	727.4209	*364.2141*	5
5	*555.3249*	278.1661	P	*571.3198*	286.1636	4
6	*684.3675*	342.6874	E	474.2671	237.6372	3
7	*854.4730*	427.7402	AcK	*345.2245*	173.1159	2
8			R	*175.1190*	88.0631	1

Figure S33 LC-MS/MS analysis of ThrRS 426-AcK. The tandem mass spectrum of the peptide (residues 420-427) LSTRPE^{AC}R from purified ThrRS 426-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red.



#	b	b++	Seq.	y	y++	#
1	115.0502	58.0287	N			7
2	244.0928	122.5500	E	763.4349	382.2211	6
3	414.1983	207.6028	AcK	634.3923	317.6998	5
4	527.2824	264.1448	I	464.2867	232.6470	4
5	584.3039	292.6556	G	351.2027	176.1050	3
6	731.3723	366.1898	F	294.1812	147.5942	2
7			K	147.1128	74.0600	1

Figure S34 LC-MS/MS analysis of ThrRS 577-AcK. The tandem mass spectrum of the peptide (residues 575-581) NEK^{AC}IGFK from purified ThrRS 577-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red.



#	b	b^{++}	Seq.	y	y^{++}	#
1	100.0757	50.5415	V			16
2	197.1285	99.0679	P	1753.8135	**877.4104**	15
3	**360.1918**	180.5995	Y	***1656.7608***	**828.8840**	14
4	**491.2323**	246.1198	M	***1493.6974***	747.3524	13
5	**604.3163**	302.6618	L	***1362.6570***	681.8321	12
6	**703.3847**	352.1960	V	***1249.5729***	625.2901	11
7	**863.4154**	432.2113	C	***1150.5045***	575.7559	10
8	920.4369	460.7221	G	***990.4738***	495.7406	9
9	**1035.4638**	518.2355	D	***933.4524***	467.2298	8
10	**1205.5693**	603.2883	AcK	***818.4254***	409.7164	7
11	**1334.6119**	667.8096	E	***648.3199***	324.6636	6
12	**1433.6803**	717.3438	V	***519.2773***	260.1423	5
13	1562.7229	781.8651	E	***420.2089***	210.6081	4
14	1649.7550	825.3811	S	***291.1663***	146.0868	3
15	1706.7764	853.8918	G	204.1343	102.5708	2
16			K	147.1128	74.0600	1

Figure S35 LC-MS/MS analysis of ThrRS 599-AcK. The tandem mass spectrum of the peptide (residues 590-605) VPYMLVCGDK^{AC}EVESGK from purified ThrRS 599-AcK. K^{AC} denotes AcK incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series. Matched peaks are in red.

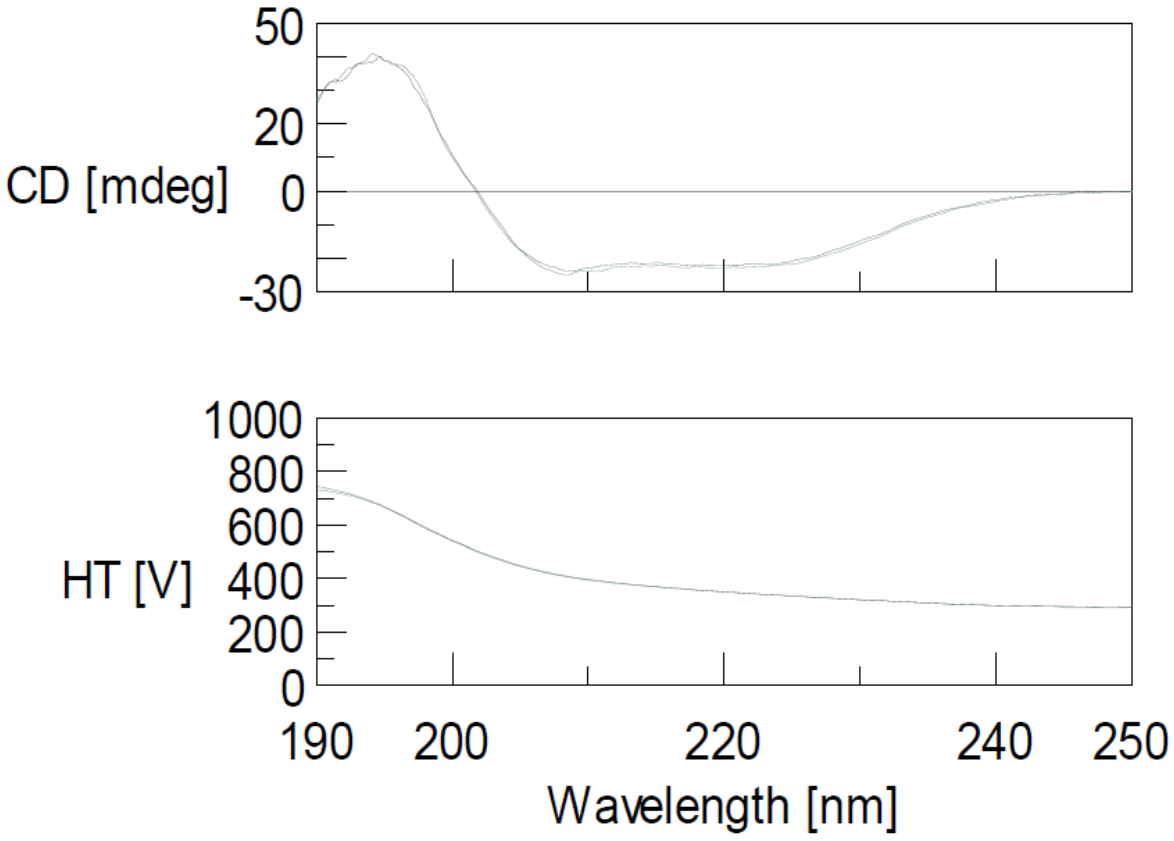


Figure S36 The circular dichroism (CD) spectra of the wild-type ThrRS (blue lines) and its 577AcK variant (green lines). The CD spectra were recorded on a J-1500 CD Spectrometer. Purified enzymes were diluted to a concentration of 0.1 mg/ml in 5 mM Tris-HCl pH 7.8, 0.1 M KCl, and scanned from 190 nm to 250 nm with a 20 nm/min speed. Scanning was performed three times for each sample and the average was plotted.

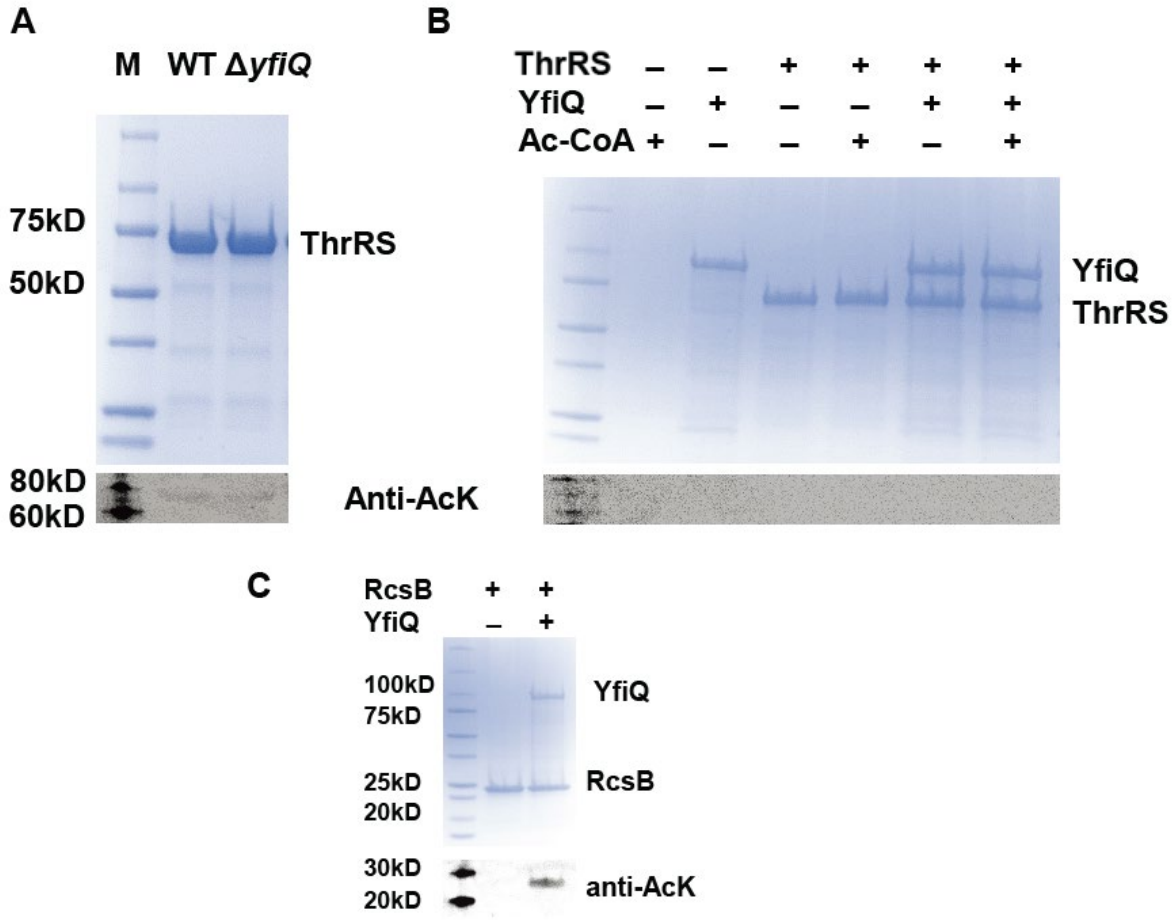


Figure 37 Acetylation of ThrRS by YfiQ. A) SDS-PAGE and western blotting analyses of purified native ThrRSs from BW25113 wild-type and Δ*yfiQ* cells. B) SDS-PAGE and western blotting analyses of purified ThrRS treated with purified YfiQ and acetyl-CoA *in vitro*. C) RcsB is a known substrate of YfiQ and was used to show the purified YfiQ protein was active.

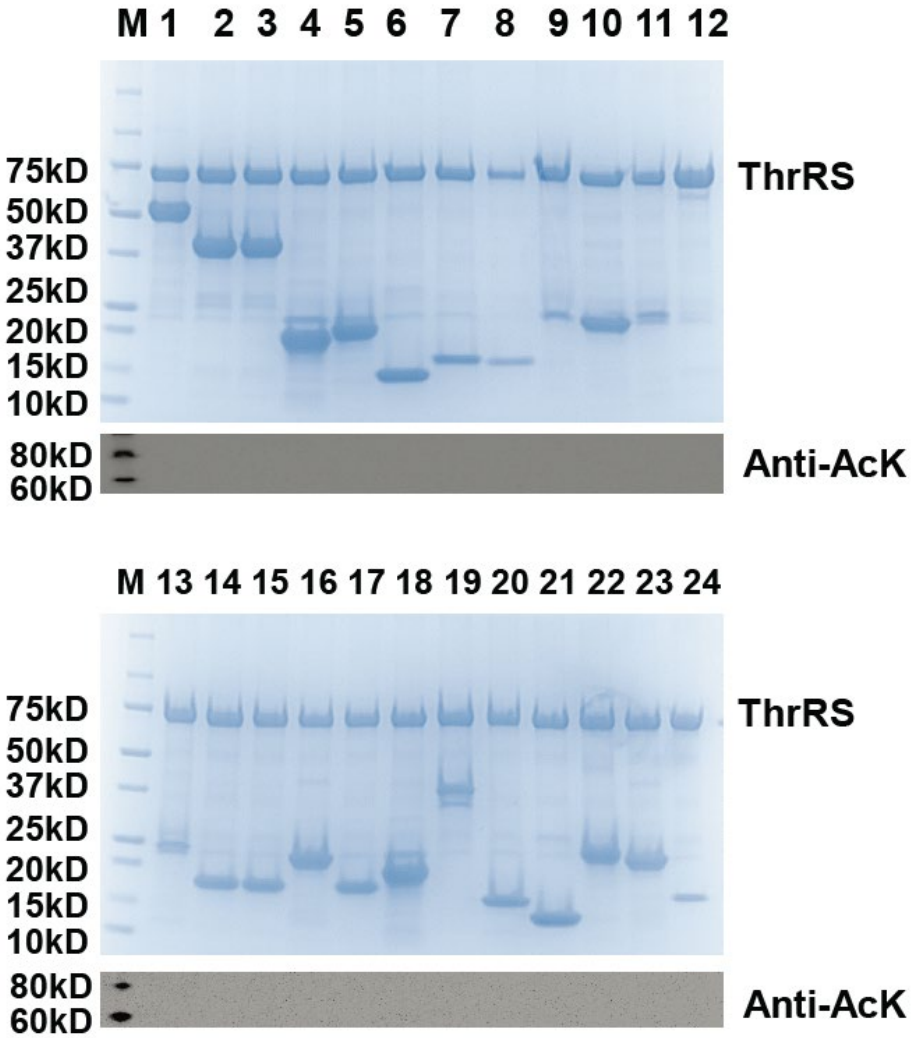


Figure S38 SDS-PAGE and western blotting of ThrRS treated with 24 GNAT family members and acetyl-CoA *in vitro*. Proteins from lane 1 to lane 24 are ArgA, AstA, CitC, ElaA, MnaT, PanM, PhnO, RimI, RimJ, RimL, SpeG, TmcA, WecD, YafP, YedL, YhbS, YhhY, YiaC, YiiD, YjaB, YjdJ,YjhQ, YjgM, and YpeA.

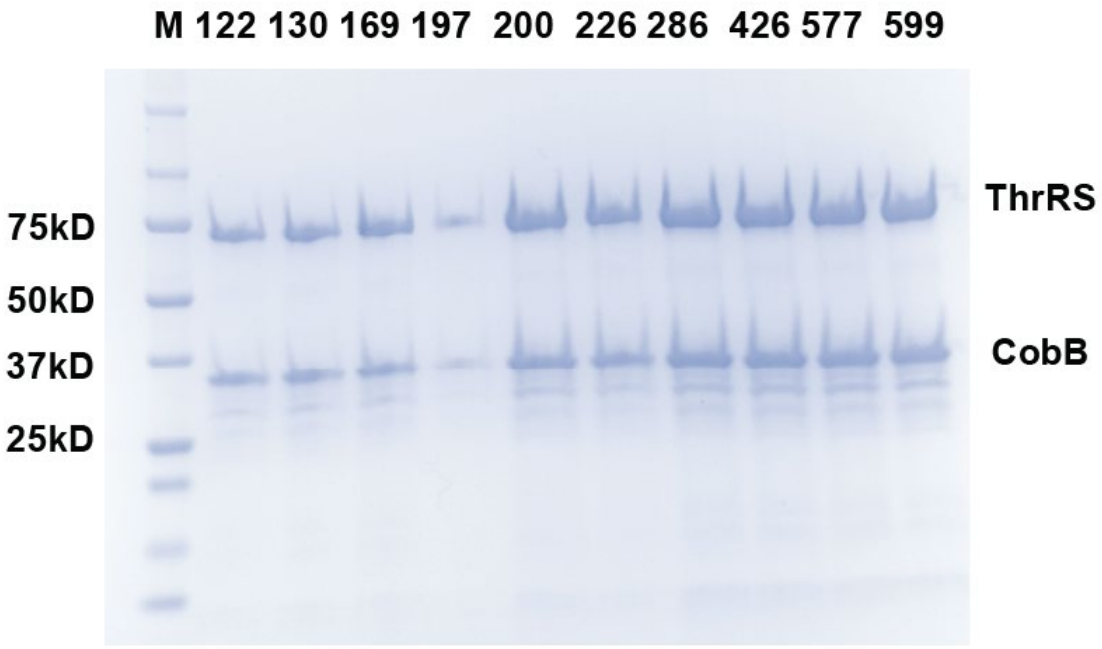


Figure S39 SDS-PAGE analyses of CobB-treated site-specifically acetylated ThrRS variants.

Supplementary data for chapter III

Supplementary material and methods

The DNA sequence of Thr-free GFP*

atgagcaagggcgaagaactgttttcgggcgtggtgccgattctggtggaactggatggtgatgtcaatggtcacaaattcagcgtgcgcg

gcgaaggtgaaggcgatgcaagcaatggtaaactgtcgctgaagtttatttgcagctcgggtaaactgccggttccgtggccgagcctggtc

agctcgctgtcgtatggtgttcagtgtttcagtcgttacccggatcacatgaaacgccacgacttttcaagtccgcgatgccggaaggttatgt

ccaagaacgtagcatctcatttaaagatgacggcagctacaaatcgcgcgccgaagtgaaattcgaaggtgattcgctggttaaccgtattga

actgaaaggcatcgattttaaggaagacggtaatattctgggccataaactggaatataacttcaattcgcacaacgtgtacatcagcgcagat

aagcagaagaacggtatcaaggctaacttcaagatccgccataatgtggaagatggcagcgttcaactggccgaccactatcagcaaaac

agcccgattggtgatggcccggtcctgctgccggacaatcattacctgagctcgcagtctgtgctgagtaaagatccgaacgaaaagcgtga

ccacatggtcctgctggaattcgtgagcgcggccggcatctcgcacggtatggacgaactgtataaaggctcataa

* The mutated Ser codons to Thr codons are marked with yellow.

The protein sequence of Thr-free GFP*

MSKGEELFSGVVPILVELDGDVNGHKFSVRGEGEGDASNGKLSLKFICSSGKLPVPWPS

LVSSLSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERSISFKDDGSYKSRAEVKFEGDS

LVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYISADKQKNGIKANFKIRHNVEDGSVQL

ADHYQQNSPIGDGPVLLPDNHYLSSQSVLSKDPNEKRDHMVLLEFVSAAGISHGMDEL

YKGS

* The substitutions of Thr with Ser are marked with yellow.

Construction of sfGFP variants.

The gene of Thr-free sfGFP was cloned into the *pCDF* vector byPCR reaction and DNA assembly by the NEBuilder® HiFi DNA Assembly Cloning Kit (New England Biolabs, Ipswich, MA, USA). The forward primer is gtttaactttaataaggagatataccatgagcaagggcgaag aactgttttcgg; the reverse primer is cagcggtggcagcagcctaggttaattatgagcctttatacagttcg-tccataccg. The primers for generating sfGFP variants with single Thr residue are listed in Table S1. The mutations were introduced by PCR reactions with the Q5 Site-Directed Mutagenesis Kit (New England Biolabs) following the manufacturer's protocol.

Protein expression and purification.

To express individual sfGFP variant, 10 mL overnight culture was transferred into 400 mL of fresh LB medium supplemented with 100 μg/mL streptomycin and grown at 37°C to an absorbance of ~ 0.6 to 0.8 at 600 nm. Then protein expression was induced by the addition of 0.1 mM IPTG. Cells were incubated at 30°C for an additional 4 h and harvested by centrifugation at 3,000 × g for 15 min at 4 °C. The cell paste was suspended in 15 mL of lysis buffer[50 mM Tris (pH 7.8), 300 mM NaCl, 20 mM imidazole, and 1 mM DTT] and broken by sonication.The crude extract was centrifuged at 20,000 × g for 20 min at 4 °C. The supernatant was filtered through a 0.45-μm membrane and loaded onto a column containing 2 mL of Ni-NTA resin (Qiagen, Hilden, Germany) previously equilibrated with 20 mL lysis buffer. The column was then washed with 20 mL of washing buffer [50 mM Tris (pH 7.8), 300 mM NaCl, and 50 mM imidazole]. The His6-tagged protein bound to the column was finally eluted with 3 mL of elution buffer [50 mM Tris(pH 7.8), 300 mM NaCl, and 150 mM imidazole].

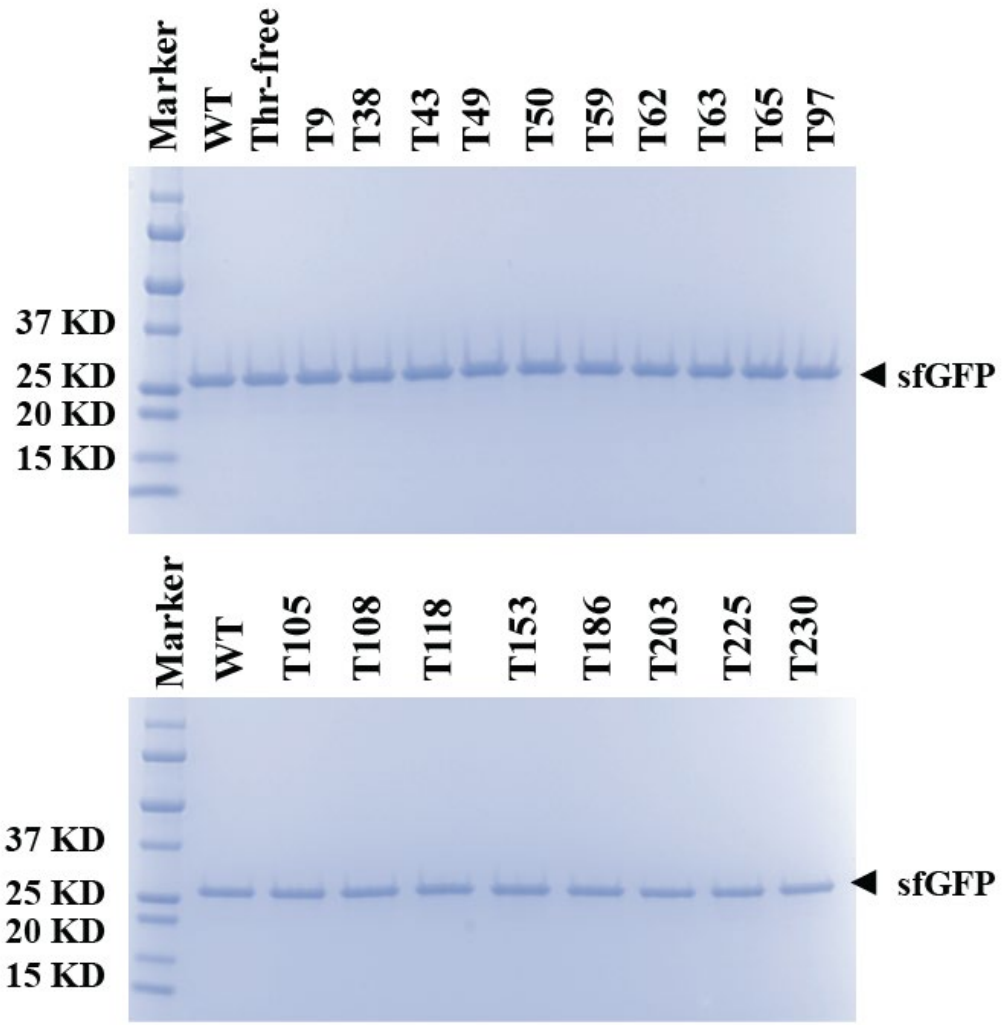


Figure S1 SDS-PAGE of purified sfGFP variants. The sample volume of elution fraction was loaded on the gel for each strain expressing sfGFP variants, respectively.

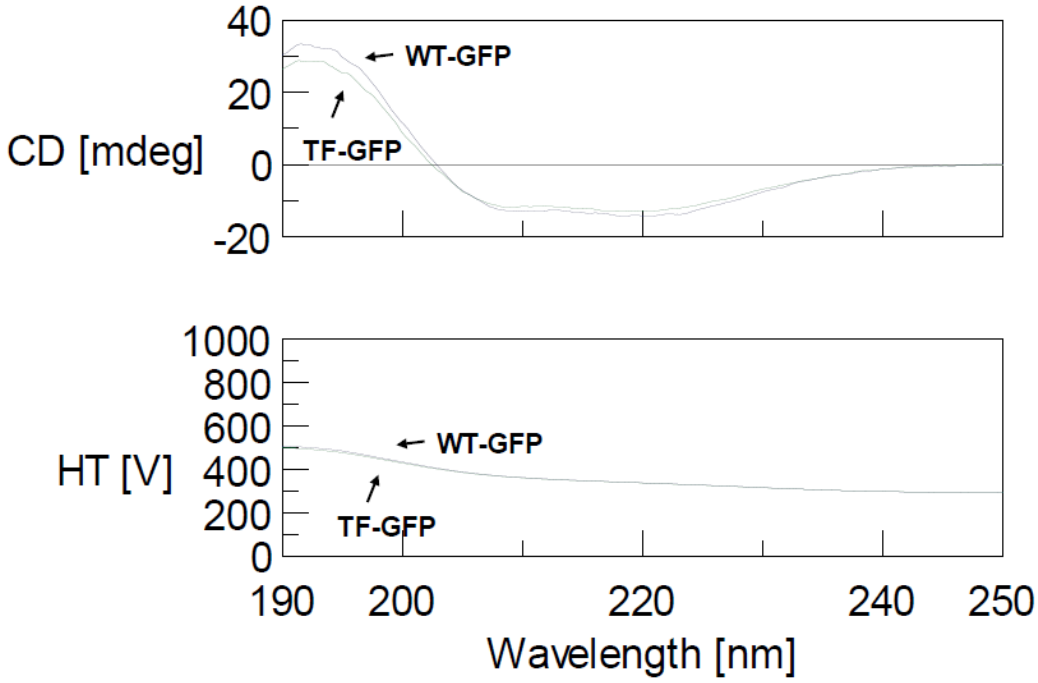


Figure S2 The circular dichroism (CD) spectra of the WT-GFP and TF-GFP. The CD spectra were recorded on a J-1500 CD Spectrometer. Purified proteins were diluted to a concentration of 0.1 mg/ml in 5 mM Tris-HCl pH 7.8, 0.1 M KCl, and scanned from 190 nm to 250 nm with a 20 nm/min speed. Scanning was performed three times for each sample and the average was plotted.

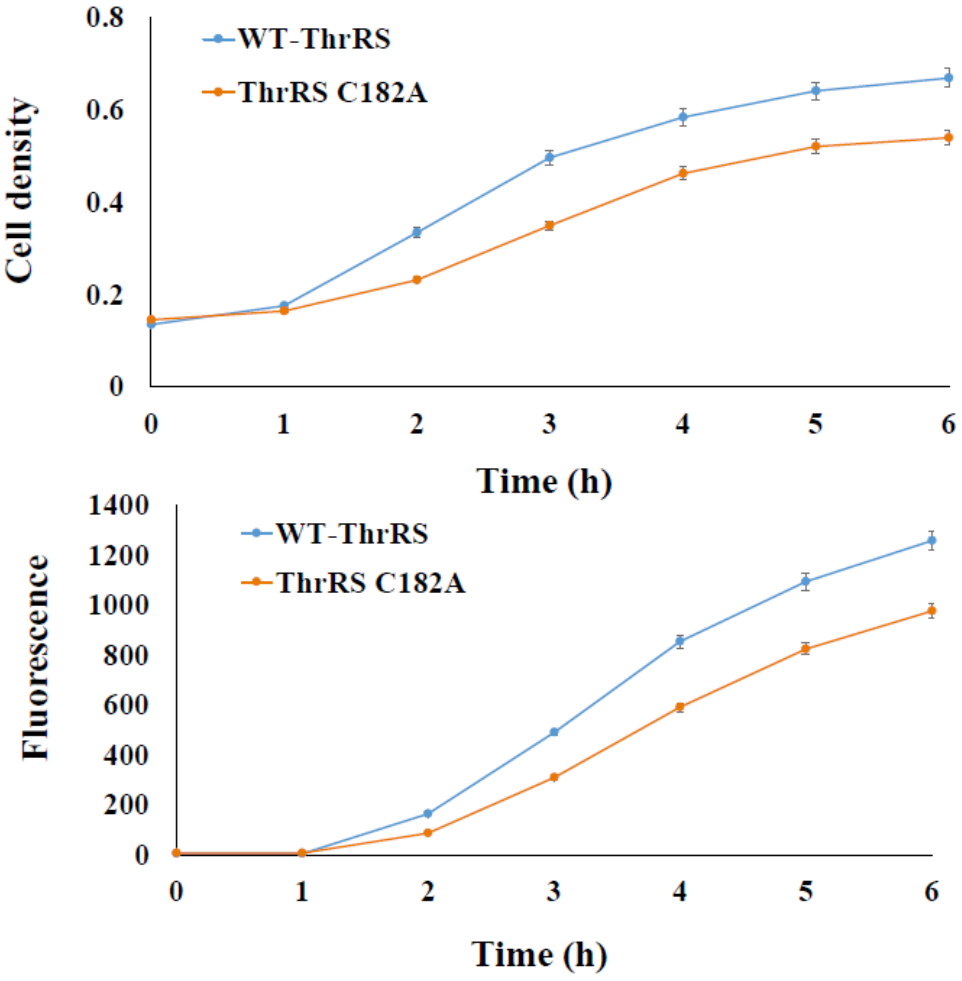


Figure S3 Cell culture densities and fluorescence readings of cells expressing the TF-sfGFP T203 reporter in strains with WT-ThrRS and ThrRS C182 variants, respectively. Mean and standard deviations were calculated based on five replicates.

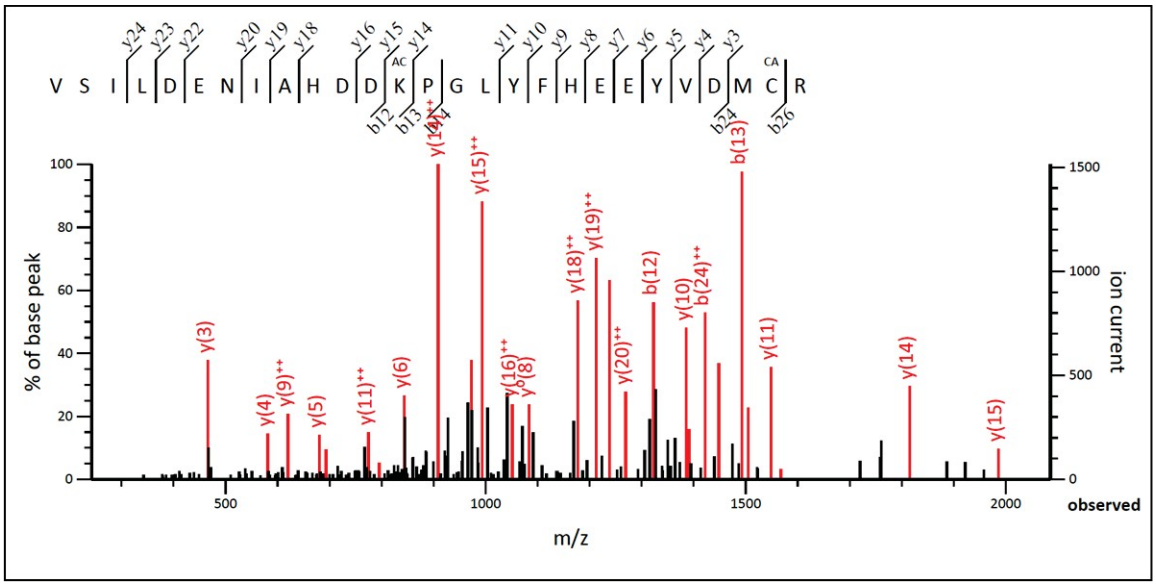


Figure S4 LC-MS/MS analysis of the ThrRS-169AcK variant. The tandem mass spectrum of the peptide (residues 157-183) VSILDENIAHDDK^{AC}PGLYFHEEYVDMCR from the purified ThrRS-169AcK variant. K^{AC} denotes AcK (acetyllysine) incorporation. The partial sequence of the peptide containing the AcK can be read from the annotated b or y ion series.

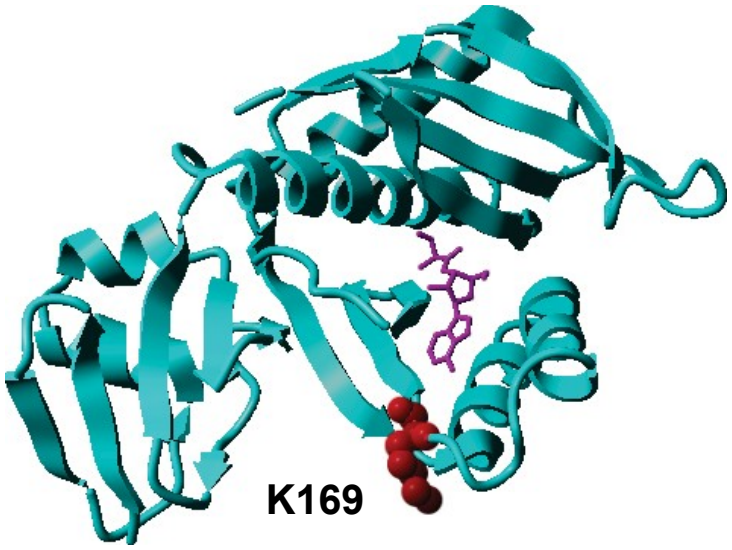


Figure S5 The location of K169 in *E. coli* ThrRS. K169 is shown in the structure of the N-terminalediting site of ThrRS (PDB ID: 1TKG). The nonhydrolyzable analog of seryl-adenylate is in magenta. The lysine residue K169 is marked with red color.

Table S1 The list of primers used for generating sfGFP variants

Variant name	Forward primer	Reverse primer
TFGFP-9T	ctgtttacgggcgtggtgcc	ttcttcgcccttgctcatggtatatc
TFGFP-38T	gatgcaaccaatggtaaactgtcg	gccttcaccttcgccgc
TFGFP-43T	gtaaactgacgctgaagtttatttgc	cattgcttgcatcgccttcac
TFGFP-49T	atttgcacctcgggtaaactgcc	aaacttcagcgacagtttaccattgcttg
TFGFP-50T	atttgcagcacgggtaaactgcc	aaacttcagcgacagtttaccattgcttg
TFGFP-59T	gccgaccctggtcagctc	cacggaaccggcagtttacc
TFGFP-62T	ggtcacctcgctgtcgtatgg	aggctcggccacggaac
TFGFP-63T	ggtcagcacgctgtcgtatgg	aggctcggccacggaac
TFGFP-65T	ggtcagctcgctgacgtatgg	aggctcggccacggaac
TFGFP-97T	gaacgtaccatctcatttaaagatgacgg	ttggacataaccttccggcatc
TFGFP-105T	gacggcacctacaaatcgcgc	atctttaaatgagatgctacgttcttggac
TFGFP-108T	gacggcagctacaaaacgcgc	atctttaaatgagatgctacgttcttggac
TFGFP-118T	gtgatacgctggttaaccgtattgaac	cttcgaatttcacttcggcgcg
TFGFP-153T	tacatcaccgcagataagcagaagaac	cacgttgtgcgaattgaagttatattccag
TFGFP-186T	caaaacaccccgattggtgatgg	ctgatagtggtcggccagttg
TFGFP-203T	ctgagcacgcagtctgtgctgag	gtaatgattgtccggcagcaggac
TFGFP-225T	gaattcgtgaccgcggcc	cagcaggaccatgtggtcac
TFGFP-230T	ggcatcacgcacggtatgg	ggccgcgctcacgaattc